

## (12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number  
**WO 01/07612 A2**

(51) International Patent Classification<sup>7</sup>: C12N 15/12,  
C07K 14/72, 16/28, G01N 33/50, 33/566, C12Q 1/68,  
A61K 38/17, 31/00

(21) International Application Number: PCT/US00/20035

(22) International Filing Date: 21 July 2000 (21.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/145,232 21 July 1999 (21.07.1999) US  
60/158,578 7 October 1999 (07.10.1999) US  
60/165,192 12 November 1999 (12.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/158,578 (CIP)
Filed on	7 October 1999 (07.10.1999)
US	60/165,192 (CIP)
Filed on	12 November 1999 (12.11.1999)
US	60/145,232 (CIP)
Filed on	21 July 1999 (21.07.1999)

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale,

CA 94087 (US). AZIMZAI, Yalda [US/US]; 2045 Rock Springs Drive, Hayward, CA 94545 (US). BURFORD, Neil [GB/US]; 1308 4th Avenue, San Francisco, CA 94122 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #12, Mountain View, CA 94040 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

## Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RECEPTORS AND ASSOCIATED PROTEINS

(57) Abstract: The invention provides human receptors and associated proteins (RECAP) and polynucleotides which identify and encode RECAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of RECAP.

WO 01/07612 A2

## RECEPTORS AND ASSOCIATED PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of receptors and associated  
5 proteins and to the use of these sequences in the diagnosis, treatment, and prevention of neurological disorders; immunological disorders, including autoimmune/inflammatory disorders; and cell proliferative disorders, including cancer.

### BACKGROUND OF THE INVENTION

10 The term receptor describes a protein that specifically recognizes other molecules. The category is broad and includes proteins with a variety of functions. The bulk of receptors are cell surface proteins which bind extracellular ligands and produce cellular responses in the areas of growth, differentiation, endocytosis, and immune response. Other receptors facilitate the selective transport of proteins out of the endoplasmic reticulum and localize enzymes to particular locations in the cell. Propagation of cellular  
15 signals, and transport and localization of proteins, all rely upon specific interactions between receptors and a variety of associated proteins. The term receptor may also be applied to proteins which bind to ligands with known or unknown chemical composition and which interact with other cellular components. For example, the steroid hormone receptors bind to and regulate transcription of DNA.

Cell surface receptors are typically integral plasma membrane proteins. These receptors  
20 recognize hormones such as catecholamines; peptide hormones; growth and differentiation factors; small peptide factors such as thyrotropin-releasing hormone; galanin, somatostatin, and tachykinins; and circulatory system-borne signaling molecules. Cell surface receptors on immune system cells recognize antigens, antibodies, and major histocompatibility complex (MHC)-bound peptides. Other cell surface receptors bind ligands to be internalized by the cell. This receptor-mediated endocytosis functions in the  
25 uptake of low density lipoproteins (LDL), transferrin, glucose- or mannose-terminal glycoproteins, galactose-terminal glycoproteins, immunoglobulins, phosphovitellogenins, fibrin, proteinase-inhibitor complexes, plasminogen activators, and thrombospondin (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York NY, p. 723; and Mikhailenko, I. et al. (1997) J. Biol. Chem. 272:6784-6791).

30 Signal transduction is the process of biochemical events by which cells are able to communicate with one another and respond to extracellular signals. Extracellular signals are transduced through a biochemical cascade that begins with the binding of a signal molecule to a cell membrane receptor. The signal is propagated to effector molecules by intracellular signal transducing proteins and culminates with the activation of an intracellular target molecule. The process of signal  
35 transduction regulates a wide variety of cell functions including cell proliferation, differentiation, and

gene transcription.

### G-protein Coupled Receptors (GPCRs)

G-protein coupled receptors (GPCRs) are a class of molecules that participate in signal transduction in a variety of cell types. GPCRs are integral membrane proteins characterized by the presence of seven hydrophobic transmembrane domains which span the plasma membrane and form a bundle of antiparallel alpha ( $\alpha$ ) helices. These proteins range in size from under 400 to over 1000 amino acids (Strosberg, A.D. (1991) *Eur. J. Biochem.* 196:1-10; Coughlin, S.R. (1994) *Curr. Opin. Cell Biol.* 6:191-197). The amino-terminus of the GPCR is extracellular, of variable length and often glycosylated; the carboxy-terminus is cytoplasmic and generally phosphorylated. Extracellular loops of the GPCR alternate with intracellular loops and link the transmembrane domains. The most conserved domains of GPCRs are the transmembrane domains and the first two cytoplasmic loops. The transmembrane domains account for structural and functional features of the receptor. In most cases, the bundle of  $\alpha$  helices forms a binding pocket. In addition, the extracellular N-terminal segment or one or more of the three extracellular loops may also participate in ligand binding. Ligand binding activates the receptor by inducing a conformational change in intracellular portions of the receptor. The activated receptor, in turn, interacts with an intracellular heterotrimeric guanine nucleotide binding (G) protein complex which mediates further intracellular signaling activities, generally the production of second messengers such as cyclic AMP (cAMP), phospholipase C, inositol triphosphate, or interactions with ion channel proteins. (Baldwin, J.M. (1994) *Curr. Opin. Cell Biol.* 6:180-190; Watson, S. and S. Arkininstall (1994) The G-protein Linked Receptor Facts Book, Academic Press, San Diego CA, pp. 2-6.) Hydrolysis of bound GTP by the G-protein completes the cycle, returning the G-protein to its inactive GDP-bound state.

GPCRs include receptors for sensory signal mediators (e.g., light and olfactory stimulatory molecules); adenosine, bombesin, bradykinin, endothelin,  $\gamma$ -aminobutyric acid (GABA), hepatocyte growth factor, luteinizing hormone (LH), thrombin, thyroid stimulating hormone (TSH), melanocortins, neuropeptide Y, opioid peptides, opsins, somatostatin, tachykinins, vasoactive intestinal polypeptide family, and vasopressin; biogenic amines (e.g., dopamine, epinephrine and norepinephrine, histamine, glutamate (metabotropic effect), acetylcholine (muscarinic effect), and serotonin); chemokines; lipid mediators of inflammation (e.g., prostaglandins and prostanoids, platelet activating factor, and leukotrienes); and peptide hormones (e.g., calcitonin, C5a anaphylatoxin, follicle-stimulating hormone (FSH), gonadotropic-releasing hormone (GnRH), neurokinin, and thyrotropin-releasing hormone (TRH), and oxytocin). GPCRs which act as receptors for stimuli that have yet to be identified are known as orphan receptors. For example, the TPRA40 protein is a GPCR isolated from mouse adipocytes and present in a number of mouse and human tissues, whose expression in adipose tissue is altered with aging and type 2 diabetes (Yang, H. (1999) *Endocrinology* 140:2859-2867).

GPCR mutations, which may cause loss of function or constitutive activation, have been associated with numerous human diseases (Coughlin, supra). For instance, retinitis pigmentosa may arise from mutations in the rhodopsin gene. Rhodopsin is the retinal photoreceptor which is located within the discs of the eye rod cell. Parma, J. et al. (1993, Nature 365:649-651) report that somatic  
5 activating mutations in the thyrotropin receptor cause hyperfunctioning thyroid adenomas and suggest that certain GPCRs susceptible to constitutive activation may behave as protooncogenes. Elevated levels of TSH receptor have been observed in brain tissue from Down syndrome and Alzheimer's disease patients, suggesting an apoptotic role for this receptor in neurodegenerative disorders (Labudova, O. et al. (1999) Life Sci. 64:1037-1044). Many clinically relevant drugs act on GPCRs, including  $\alpha$  and  $\beta$   
10 blockers which affect the activity of adrenergic receptors and are used in the treatment of hypertension and other cardiovascular disorders (Watson, supra, pp. 32-33).

#### Receptors Involved in the Immune System

Examples of GPCRs implicated in inflammation and the immune response include the EGF  
15 module-containing, mucin-like hormone receptor (Emr1) and CD97 receptor proteins. These GPCRs are members of the recently characterized EGF-TM7 receptors family. These seven transmembrane hormone receptors exist as heterodimers in vivo and contain between three and seven potential calcium-binding EGF-like motifs. CD97 is predominantly expressed in leukocytes, and is markedly upregulated on activated B and T cells. (McKnight, A. J. and Gordon, S. (1998) J. Leukoc. Biol.  
20 63:271-280.)

Irregularities in the GPCR signaling cascade may result in abnormal activation of leukocytes and lymphocytes, leading to the tissue damage and destruction seen in many inflammatory and autoimmune diseases such as rheumatoid arthritis, biliary cirrhosis, hemolytic anemia, lupus erythematosus, and thyroiditis. Abnormal cell proliferation, including cyclic AMP stimulation of brain, thyroid, adrenal, and  
25 gonadal tissue proliferation is regulated by G proteins (Meij, J.T.A. (1996) Mol. Cell. Biochem. 157:31-38; Aussel, C. et al. (1988) J. Immunol. 140:215-220).

T cells play a dual role in the immune system as effectors and regulators, coupling antigen recognition with the transmission of signals that induce cell death in infected cells and stimulate other immune cells. Although T cells collectively recognize a wide range of different antigens, a clonal line of  
30 T cells can only recognize a single antigen. Moreover, the antigen must be presented to the T cell receptor (TCR) as a peptide complexed with a major histocompatibility molecule (MHC) on the surface of an antigen-presenting cell. The TCR on most T cells consists of two polypeptide subunits,  $\alpha$  and  $\beta$ , which are immunoglobulin-like integral membrane glycoproteins of similar molecular weight. The TCR $\alpha$  and TCR $\beta$  subunits have an extracellular domain containing both variable and constant regions, a  
35 transmembrane domain that traverses the membrane once, and a short intracellular domain (Saito, H. et

al. (1984) *Nature* 309:757-762). The genes for the TCR subunits are constructed through somatic rearrangement of different gene segments. Interaction of antigen in the proper MHC context with the TCR initiates signaling cascades that induce the proliferation, maturation, and function of cellular components of the immune system (Weiss, A. (1991) *Annu. Rev. Genet.* 25: 487-510). Rearrangements  
 5 in TCR genes and alterations in TCR expression have been noted in lymphomas, leukemias, autoimmune disorders, and immunodeficiency disorders (Aisenberg, A.C. et al. (1985) *N. Engl. J. Med.* 313:529-533; Olive, C. (1995) *Immunol. Cell. Biol.* 73:297-307; and Weiss, *supra*). Immunizations with peptides derived from TCRs are effective treatment for some human T-cell-mediated autoimmune disease and in animal models of such illnesses, in particular, rheumatoid arthritis (Bridges, S.L. and Moreland, L.W.  
 10 (1998) *Rheum. Dis. Clin. North Am.* 24:641-650).

Tumor necrosis factor (TNF) is a pleiotropic cytokine that mediates immune regulation and inflammatory responses. The cellular responses triggered by TNF are initiated through its interaction with two distinct cell surface receptors, TNF-R1 and TNF-R2. (Tartaglia, L.A. and Goeddel, D.V. (1992) *Immunol. Today* 13:151-153). Both TNF receptors are part of the TNF receptor (TNFR)  
 15 superfamily, whose members include the Fas antigen, the p75 subunit of the NGF receptor, the TRAIL receptor, TRUNND, SalF19R, CD27, CD30, and CD40. Members of the TNFR superfamily share the TNFR/NGFR family cysteine-rich region signature, which consists of cysteine-rich pseudo-repeats in the extracellular domains. (ExPASy PROSITE document PDOC00561; Pan, G. et al. (1998) *FEBS Lett.* 424:41-45; Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221; and Smith, C.A. et al. (1994) *Cell*  
 20 76:959-962). Polymorphisms in TNF-R2 are associated with systemic lupus erythematosus (Komata, T. et al. (1999) *Tissue Antigens* 53:527-533). In addition, increased serum concentrations of soluble TNF-R1 have been observed in some patients with advanced gastric or colorectal cancer (Shibata, M. et al. (1998) *Surg. Today* 28:884-888).

Another essential component of the immune response is the complement system, which responds  
 25 to signals provided by antigen recognition by mobilizing effector activities including inflammation, phagocytosis, and cell lysis. Receptors on macrophages and neutrophils bind activated complement C3 on the surface of foreign particles such as bacteria, thus targeting the foreign particles for phagocytosis and destruction by lysosomal enzymes. Complement receptor 1 (CR1) has a wide cellular/tissue distribution, and mediates enhancement of phagocytosis, induction of IL-1 secretion and enhancement of  
 30 B-cell differentiation. Defective expression of CR1 is associated with the autoimmune disease systemic lupus erythematosus. (Carroll, M.C. (1998) *Annu. Rev. Immunol.* 16:545-568.)

#### Nuclear Receptors

The nuclear receptors are another receptor family, and includes the retinoic acid receptors  
 35 (RARs) and the retinoid X receptors (RXRs). RARs and RXRs can form heterodimers which are

thought to have a signal transduction function. Retinoic acid (RA) is a biologically active metabolite of vitamin A (retinol), a fat-soluble vitamin found mainly in fish liver oils, liver, egg yolk, butter, and cream. Retinol cannot be synthesized in vivo and must be obtained from the diet. Retinol, RA, and other retinoids influence epithelial cell differentiation. A number of carrier proteins which bind retinol or other  
5 retinoids have been identified. These retinoid binding proteins (RBPs) appear to direct bound retinoid molecules to specific metabolic pathways. Specific receptors for RBPs mediate the cellular uptake of retinoids and the transfer of retinoids to intracellular RBPs (Sundaram, M. et al. (1999) J. Biol. Chem. 273:3336-3342).

#### 10 Low Molecular Weight (LMW) G-proteins

Low molecular weight (LMW) G-proteins regulate cell growth, cell cycle control, protein secretion, and intracellular vesicle interaction. They consist of single polypeptides which are able to bind to and hydrolyze GTP, thus cycling between an inactive and an active state. LMW G-proteins respond to extracellular signals from receptors and activating proteins by transducing mitogenic signals involved in  
15 various cell functions. The binding and hydrolysis of GTP regulates the response of LMW G-proteins and acts as an energy source during this process (Bokoch, G. M. and Der, C. J. (1993) FASEB J. 7:750-759).

At least sixty members of the LMW G-protein superfamily have been identified and are currently grouped into the ras, rho, arf, sar1, ran, and rab subfamilies. Activated ras genes were initially found in  
20 human cancers and subsequent studies confirmed that ras function is critical to receptor tyrosine kinase-mediated signal transduction pathways that determine whether cells continue to grow and divide, or whether they differentiate. Rho G-proteins control signal transduction pathways that link growth factor receptors to actin polymerization, which is necessary for normal cellular growth and division. The rab, arf, and sar1 families of proteins control the translocation of vesicles to and from membranes for protein  
25 localization, protein processing, and secretion. Ran G-proteins are located in the nucleus and have a key role in nuclear protein import, the control of DNA synthesis, and cell-cycle progression (Hall, A. (1990) Science 249:635-640; Barbacid, M. (1987) Ann. Rev Biochem. 56:779-827; and Sasaki, T. and Takai, Y. (1998) Biochem. Biophys. Res. Commun. 245:641-645).

LMW G-proteins are GTPases which cycle between the active GTP-bound and inactive  
30 GDP-bound forms. At least three types of proteins regulate this process: GTPase-activating proteins, (GAP), which stimulate GTP hydrolysis by the LMW G-protein; guanine nucleotide exchange factors (GEP), which facilitate the exchange of GDP bound to the LMW G-protein for GTP; and guanine nucleotide dissociation inhibitors (GDI), which inhibit this reaction (Ikeda, M. et al. (1998) J. Biol. Chem. 273:814-821; Quilliam, L. A. (1995) Bioessays 17:395-404). The best characterized GEP is the  
35 mammalian homologue of the Drosophila Son-of-Sevenless protein. Both GEP and GAP activity may be

affected by extracellular stimuli and modified by accessory proteins such as RalBP1 and POB1. Mutant Ras-family proteins, which bind but can not hydrolyze GTP, are permanently activated, and cause cell proliferation or cancer, as do GEP that activate LMW G-proteins (Drivas, G. T. et al. (1990) Mol. Cell. Biol. 10:1793-1798; and Whitehead, I. P. et al. (1998) Mol Cell Biol. 18:4689-4697).

5

#### Olfactory GPCRs

Another large subfamily of GPCRs are the olfactory receptors. These receptors share the seven hydrophobic transmembrane domains of other GPCRs and function by registering G protein-mediated transduction of odorant signals. Numerous distinct olfactory receptors are required to distinguish  
10 different odors. Each olfactory sensory neuron expresses only one type of olfactory receptor, and distinct spatial zones of neurons expressing distinct receptors are found in nasal passages.

The discovery of new receptors and associated proteins and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative, autoimmune/inflammatory, and neurological disorders.

15

### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, receptors and associated proteins, referred to collectively as "RECAP" and individually as "RECAP-1," "RECAP-2," "RECAP-3," "RECAP-4," "RECAP-5," "RECAP-6," "RECAP-7," "RECAP-8," "RECAP-9," "RECAP-10," "RECAP-11,"  
20 "RECAP-12," "RECAP-13," "RECAP-14," "RECAP-15," "RECAP-16," "RECAP-17," "RECAP-18," "RECAP-19," "RECAP-20," "RECAP-21," and "RECAP-22." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the  
25 group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-22.

The invention further provides an isolated polynucleotide encoding a polypeptide comprising an  
30 amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting  
35 of SEQ ID NO:1-22. In one alternative, the polynucleotide encodes a polypeptide selected from the

group consisting of SEQ ID NO:1-22. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:23-44.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence  
5 selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-  
10 22. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group  
15 consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) culturing a cell under conditions suitable for expression  
20 of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid  
25 sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22.

30 The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)  
35 an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous



nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of  
5 SEQ ID NO:23-44, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the  
10 sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said  
15 target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e)  
20 an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a pharmaceutical composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid  
25 sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and a pharmaceutically acceptable excipient. In  
30 one embodiment, the pharmaceutical composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional RECAP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

The invention also provides a method for screening a compound for effectiveness as an agonist  
35 of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino

acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence  
5 selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a pharmaceutical composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of  
10 functional RECAP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring  
15 amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in  
20 the sample. In one alternative, the invention provides a pharmaceutical composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional RECAP, comprising administering to a patient in need of such treatment the pharmaceutical composition.

25 The invention further provides a method of screening for a compound that specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from  
30 the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

35 The invention further provides a method of screening for a compound that modulates the

activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-22. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:23-44, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, ii) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23-44, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of the above polynucleotide sequence; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization

complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 shows polypeptide and nucleotide sequence identification numbers (SEQ ID NOs), clone  
5 identification numbers (clone IDs), cDNA libraries, and cDNA fragments used to assemble full-length  
sequences encoding RECAP.

Table 2 shows features of each polypeptide sequence, including potential motifs, homologous  
sequences, and methods, algorithms; and searchable databases used for analysis of RECAP.

Table 3 shows selected fragments of each nucleic acid sequence; the tissue-specific expression  
10 patterns of each nucleic acid sequence as determined by northern analysis; diseases, disorders, or  
conditions associated with these tissues; and the vector into which each cDNA was cloned.

Table 4 describes the tissues used to construct the cDNA libraries from which cDNA clones  
encoding RECAP were isolated.

Table 5 shows the tools, programs, and algorithms used to analyze the polynucleotides and  
15 polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood  
that this invention is not limited to the particular machines, materials and methods described, as these  
20 may vary. It is also to be understood that the terminology used herein is for the purpose of describing  
particular embodiments only, and is not intended to limit the scope of the present invention which will be  
limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an,"  
and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a  
25 reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a  
reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings  
as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any  
machines, materials, and methods similar or equivalent to those described herein can be used to practice  
30 or test the present invention, the preferred machines, materials and methods are now described. All  
publications mentioned herein are cited for the purpose of describing and disclosing the cell lines,  
protocols, reagents and vectors which are reported in the publications and which might be used in  
connection with the invention. Nothing herein is to be construed as an admission that the invention is not  
entitled to antedate such disclosure by virtue of prior invention.

**DEFINITIONS**

"RECAP" refers to the amino acid sequences of substantially purified RECAP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

5       The term "agonist" refers to a molecule which intensifies or mimics the biological activity of RECAP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of RECAP either by directly interacting with RECAP or by acting on components of the biological pathway in which RECAP participates.

      An "allelic variant" is an alternative form of the gene encoding RECAP. Allelic variants may  
10 result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a  
15 given sequence.

      "Altered" nucleic acid sequences encoding RECAP include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as RECAP or a polypeptide with at least one functional characteristic of RECAP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of  
20 the polynucleotide encoding RECAP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding RECAP. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent RECAP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility,  
25 hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of RECAP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains  
30 having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

      The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein  
35 molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the

complete native amino acid sequence associated with the recited protein molecule.

“Amplification” relates to the production of additional copies of a nucleic acid sequence.

Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

5       The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of RECAP. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of RECAP either by directly interacting with RECAP or by acting on components of the biological pathway in which RECAP participates.

10       The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind RECAP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or  
15 synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

      The term “antigenic determinant” refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host  
20 animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

      The term “antisense” refers to any composition capable of base-pairing with the “sense” (coding)  
25 strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by  
30 any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation “negative” or “minus” can refer to the antisense strand, and the designation “positive” or “plus” can refer to the sense strand of a reference DNA molecule.

35       The term “biologically active” refers to a protein having structural, regulatory, or biochemical

functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic RECAP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution.

10 Compositions comprising polynucleotide sequences encoding RECAP or fragments of RECAP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

15 "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (PE Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University  
20 of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino  
25 acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
30	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
35	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val

5	Lys Met Phe Ser Thr Trp Tyr Val	Arg, Gln, Glu Leu, Ile His, Met, Leu, Trp, Tyr Cys, Thr Ser, Val Phe, Tyr His, Phe, Trp Ile, Leu, Thr
---	--	--

---

10       Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

      A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the  
15   absence of one or more amino acid residues or nucleotides.

      The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one  
20   modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

      A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

      A "fragment" is a unique portion of RECAP or the polynucleotide encoding RECAP which is  
25   identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid  
30   residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50% of a polypeptide) as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present  
35   embodiments.

      A fragment of SEQ ID NO:23-44 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:23-44, for example, as distinct from any other sequence in the



genome from which the fragment was obtained. A fragment of SEQ ID NO:23-44 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:23-44 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:23-44 and the region of SEQ ID NO:23-44 to which the fragment corresponds are routinely  
5 determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-22 is encoded by a fragment of SEQ ID NO:23-44. A fragment of SEQ ID NO:1-22 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-22. For example, a fragment of SEQ ID NO:1-22 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-22. The precise length of a  
10 fragment of SEQ ID NO:1-22 and the region of SEQ ID NO:1-22 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A "full-length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full-  
15 length" polynucleotide sequence encodes a "full-length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a  
20 standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence  
25 alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as  
30 the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several  
35 sources, including the NCBI, Bethesda, MD, and on the Internet at

<http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including “blastn,” that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called “BLAST 2 Sequences” that is used for direct pairwise comparison of two nucleotide sequences. “BLAST 2 Sequences” can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The “BLAST 2 Sequences” tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

10        *Matrix: BLOSUM62*  
           *Reward for match: 1*  
           *Penalty for mismatch: -2*  
           *Open Gap: 5 and Extension Gap: 2 penalties*  
           *Gap x drop-off: 50*  
 15        *Expect: 10*  
           *Word Size: 11*  
           *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

25        Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

35        Percent identity between polypeptide sequences may be determined using the default parameters

of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with  
 5 polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (Apr-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

10        *Matrix: BLOSUM62*  
           *Open Gap: 11 and Extension Gap: 1 penalties*  
           *Gap x drop-off: 50*  
           *Expect: 10*  
           *Word Size: 3*  
 15        *Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150  
 20 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for chromosome  
 25 replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a  
 30 complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding  
 35 between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing

of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100  
5 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target  
10 sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention  
15 include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide  
20 at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

25 The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_0t$  or  $R_0t$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or  
30 their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of  
35 various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and

systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of RECAP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of RECAP  
5 which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

10 The term "modulate" refers to a change in the activity of RECAP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of RECAP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or  
15 synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding  
20 sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially  
25 bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an RECAP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by  
30 cell type depending on the enzymatic milieu of RECAP.

"Probe" refers to nucleic acid sequences encoding RECAP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic  
35 acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by

complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous  
5 nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

10 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al., 1987, Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al., 1990, PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a  
15 known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000  
20 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program  
25 (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public  
30 from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful  
35 in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or

specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing nucleic acids encoding RECAP, or fragments thereof, or RECAP itself, may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the

amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally  
5 associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and  
10 capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient  
15 cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells  
20 includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid  
25 introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic  
30 organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants, and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

35 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at



least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 95% or at least 98% or greater

5 sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species

10 variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides generally will have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for

15 example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least

20 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human receptors and associated proteins (RECAP), the polynucleotides encoding RECAP, and the use of these compositions for the diagnosis,

25 treatment, or prevention of neurological disorders; immunological disorders, including autoimmune/inflammatory disorders; and cell proliferative disorders, including cancer.

Table 1 lists the Incyte clones used to assemble full length nucleotide sequences encoding RECAP. Columns 1 and 2 show the sequence identification numbers (SEQ ID NOs) of the polypeptide and nucleotide sequences, respectively. Column 3 shows the clone IDs of the Incyte clones in which

30 nucleic acids encoding each RECAP were identified, and column 4 shows the cDNA libraries from which these clones were isolated. Column 5 shows Incyte clones and their corresponding cDNA libraries. Clones for which cDNA libraries are not indicated were derived from pooled cDNA libraries. In some cases, GenBank sequence identifiers are also shown in column 5. The Incyte clones and GenBank cDNA sequences, where indicated, in column 5 were used to assemble the consensus nucleotide sequence of each

35 RECAP and are useful as fragments in hybridization technologies.

The columns of Table 2 show various properties of each of the polypeptides of the invention: column 1 references the SEQ ID NO; column 2 shows the number of amino acid residues in each polypeptide; column 3 shows potential phosphorylation sites; column 4 shows potential glycosylation sites; column 5 shows the amino acid residues comprising signature sequences and motifs; column 6 shows homologous sequences as identified by BLAST analysis along with relevant citations, all of which are expressly incorporated by reference herein in their entirety; and column 7 shows analytical methods and in some cases, searchable databases to which the analytical methods were applied. The methods of column 7 were used to characterize each polypeptide through sequence homology and protein motifs.

The columns of Table 3 show the tissue-specificity and diseases, disorders, or conditions associated with nucleotide sequences encoding RECAP. The first column of Table 3 lists the nucleotide SEQ ID NOs. Column 2 lists fragments of the nucleotide sequences of column 1. These fragments are useful, for example, in hybridization or amplification technologies to identify SEQ ID NO:23-44 and to distinguish between SEQ ID NO:23-44 and related polynucleotide sequences. The polypeptides encoded by these fragments are useful, for example, as immunogenic peptides. Column 3 lists tissue categories which express RECAP as a fraction of total tissues expressing RECAP. Column 4 lists diseases, disorders, or conditions associated with those tissues expressing RECAP as a fraction of total tissues expressing RECAP. Column 5 lists the vectors used to subclone each cDNA library. Of particular interest is the expression of SEQ ID NO:11 in hematopoietic/immune tissues and the expression of SEQ ID NO:14 in reproductive tissues.

The columns of Table 4 show descriptions of the tissues used to construct the cDNA libraries from which cDNA clones encoding RECAP were isolated. Column 1 references the nucleotide SEQ ID NOs, column 2 shows the cDNA libraries from which these clones were isolated, and column 3 shows the tissue origins and other descriptive information relevant to the cDNA libraries in column 2.

The invention also encompasses RECAP variants. A preferred RECAP variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the RECAP amino acid sequence, and which contains at least one functional or structural characteristic of RECAP.

The invention also encompasses polynucleotides which encode RECAP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44, which encodes RECAP. The polynucleotide sequences of SEQ ID NO:23-44, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding RECAP. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least

about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding RECAP. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:23-44 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:23-44. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of RECAP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding RECAP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring RECAP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode RECAP and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring RECAP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding RECAP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding RECAP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode RECAP and RECAP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding RECAP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:23-44 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the

embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (PE Biosystems, Foster City CA), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification  
5 system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (PE Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (PE Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or  
10 other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding RECAP may be extended utilizing a partial nucleotide  
15 sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a  
20 circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to  
25 insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-  
30 based methods, primers may be designed using commercially available software, such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full-length cDNAs, it is preferable to use libraries that have been  
35 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, PE Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode RECAP may be cloned in recombinant DNA molecules that direct expression of RECAP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express RECAP.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter RECAP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Cramer, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of RECAP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be

recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

5 In another embodiment, sequences encoding RECAP may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser.* 7:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser.* 7:225-232.) Alternatively, RECAP itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton,  
10 T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (PE Biosystems). Additionally, the amino acid sequence of RECAP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally  
15 occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

20 In order to express a biologically active RECAP, the nucleotide sequences encoding RECAP or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding  
25 RECAP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding RECAP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding RECAP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be  
30 needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.*  
35 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding RECAP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding RECAP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al., (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding RECAP. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding RECAP can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPT1 plasmid (Life Technologies). Ligation of sequences encoding RECAP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions

in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of RECAP are needed, e.g. for the production of antibodies, vectors which direct high level expression of RECAP may be used. For example, vectors containing the strong, inducible T5 or T7 bacteriophage promoter may be used.

5 Yeast expression systems may be used for production of RECAP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, 10 supra; and Scorer, supra.)

Plant systems may also be used for expression of RECAP. Transcription of sequences encoding RECAP may be driven viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be 15 used. (See, e.g., Coruzzi, supra; Broglie, supra; and Winter, supra.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding RECAP may be ligated into an 20 adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses RECAP in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based 25 vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

30 For long term production of recombinant proteins in mammalian systems, stable expression of RECAP in cell lines is preferred. For example, sequences encoding RECAP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before 35 being switched to selective media. The purpose of the selectable marker is to confer resistance to a



selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> and *apr*<sup>-</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding RECAP is inserted within a marker gene sequence, transformed cells containing sequences encoding RECAP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding RECAP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding RECAP and that express RECAP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of RECAP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on RECAP is preferred, but a competitive binding assay may be

employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

5 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding RECAP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding RECAP, or any fragments thereof, may be cloned into a vector for the production of  
10 an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include  
15 radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding RECAP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the  
20 vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode RECAP may be designed to contain signal sequences which direct secretion of RECAP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the  
25 polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection  
30 (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding RECAP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric RECAP protein containing a  
35 heterologous moiety that can be recognized by a commercially available antibody may facilitate the

screening of peptide libraries for inhibitors of RECAP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the RECAP encoding sequence and the heterologous protein sequence, so that RECAP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled RECAP may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

RECAP of the present invention or fragments thereof may be used to screen for compounds that specifically bind to RECAP. At least one and up to a plurality of test compounds may be screened for specific binding to RECAP. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of RECAP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which RECAP binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express RECAP, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing RECAP or cell membrane fractions which contain RECAP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either RECAP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the

assay may comprise the steps of combining at least one test compound with RECAP, either in solution or affixed to a solid support, and detecting the binding of RECAP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor.

Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural

5 product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

RECAP of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of RECAP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for RECAP activity, wherein RECAP is combined with at least one test compound, and the activity of RECAP in

10 the presence of a test compound is compared with the activity of RECAP in the absence of the test compound. A change in the activity of RECAP in the presence of the test compound is indicative of a compound that modulates the activity of RECAP. Alternatively, a test compound is combined with an in vitro or cell-free system comprising RECAP under conditions suitable for RECAP activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of RECAP  
15 may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding RECAP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of  
20 human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by

25 homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).

Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the

30 resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains.

Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding RECAP may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate

35 into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al.

(1998) Science 282:1145-1147).

Polynucleotides encoding RECAP can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding RECAP is injected into animal ES cells, and the injected sequence  
 5 integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress RECAP, e.g., by secreting RECAP in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## 10 THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of RECAP and receptors and associated proteins. In addition, the expression of RECAP is closely associated with cell proliferation, cancer, inflammation/trauma, and with neurological disorders. Therefore, RECAP appears to play a role in neurological disorders; immunological disorders, including  
 15 autoimmune/inflammatory disorders; and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased RECAP expression or activity, it is desirable to decrease the expression or activity of RECAP. In the treatment of disorders associated with decreased RECAP expression or activity, it is desirable to increase the expression or activity of RECAP.

Therefore, in one embodiment, RECAP or a fragment or derivative thereof may be  
 20 administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RECAP. Examples of such disorders include, but are not limited to, a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, Down's syndrome, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural  
 25 muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis,  
 30 tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia  
 35 gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders;

seasonal affective disorder (SAD); akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an immunological disorder, including autoimmune/inflammatory disorders, such as acquired immunodeficiency syndrome

5 (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease, Addison's disease,  
 10 adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis,  
 15 glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner  
 20 syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, hematopoietic cancers, including lymphoma, leukemia, and myeloma, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary  
 25 thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

30 In another embodiment, a vector capable of expressing RECAP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RECAP including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified RECAP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or  
 35 prevent a disorder associated with decreased expression or activity of RECAP including, but not limited

to, those provided above.

In still another embodiment, an agonist which modulates the activity of RECAP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of RECAP including, but not limited to, those listed above.

- 5 In a further embodiment, an antagonist of RECAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of RECAP. Examples of such disorders include, but are not limited to, those neurological disorders; immunological disorders, including autoimmune/inflammatory disorders; and cell proliferative disorders, including cancer, described above. In one aspect, an antibody which specifically binds RECAP may be used directly as an antagonist or  
10 indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express RECAP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding RECAP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of RECAP including, but not limited to, those described above.

- 15 In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders  
20 described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

- An antagonist of RECAP may be produced using methods which are generally known in the art. In particular, purified RECAP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind RECAP. Antibodies to RECAP may also be generated  
25 using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

- For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and  
30 others may be immunized by injection with RECAP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli  
35 Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to RECAP have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of RECAP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to RECAP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce RECAP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for RECAP may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between RECAP and its specific



antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering RECAP epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for RECAP. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of RECAP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple RECAP epitopes, represents the average affinity, or avidity, of the antibodies for RECAP. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular RECAP epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the RECAP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of RECAP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of RECAP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al., supra.)

In another embodiment of the invention, the polynucleotides encoding RECAP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding RECAP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding RECAP. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g.,

Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other  
 5 gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding RECAP may be used for  
 10 somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined  
 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic  
 15 fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial  
 hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and Somia, N. (1997) *Nature* 389:239-242)), (ii) express a  
 conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell  
 20 proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA.* 93:11395-11399), hepatitis B or C  
 virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and  
 protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a  
 25 genetic deficiency in RECAP expression or regulation causes disease, the expression of RECAP from an  
 appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic  
 deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in RECAP  
 are treated by constructing mammalian expression vectors encoding RECAP and introducing these  
 30 vectors by mechanical means into RECAP-deficient cells. Mechanical transfer technologies for use with  
 cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold  
 particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the  
 use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217;  
 Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-  
 35 450).

Expression vectors that may be effective for the expression of RECAP include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). RECAP may be expressed using (i) a

5 constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) *Proc. Natl. Acad. Sci. USA* 89:5547-5551; Gossen, M. et al. (1995) *Science* 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) *Curr. Opin. Biotechnol.* 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter

10 (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding RECAP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID

15 TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized

20 mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to RECAP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding RECAP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive

25 element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target

30 cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus

35 packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors,

transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; 5 Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding RECAP to cells which have one or more genetic abnormalities with respect to the expression of RECAP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be 10 versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544; and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by 15 reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding RECAP to target cells which have one or more genetic abnormalities with respect to the expression of RECAP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing RECAP to cells of the central nervous system, for which HSV has a 20 tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. 25 U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby 30 incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to 35 deliver polynucleotides encoding RECAP to target cells. The biology of the prototypic alphavirus,

Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotech. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full-length genomic RNA, resulting in the  
5 overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for RECAP into the alphavirus genome in place of the capsid-coding region results in the production of a large number of RECAP-coding RNAs and the synthesis of high levels of RECAP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster  
10 normal kidney cells (BHK-21).with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of RECAP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses,  
15 performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes  
20 inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript  
25 from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage  
30 of sequences encoding RECAP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural  
35 features which may render the oligonucleotide inoperable. The suitability of candidate targets may also

be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for  
5 chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding RECAP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or  
10 tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in  
15 all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding RECAP. Compounds  
20 which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide  
25 expression. Thus, in the treatment of disorders associated with increased RECAP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding RECAP may be therapeutically useful, and in the treatment of disorders associated with decreased RECAP expression or activity, a compound which specifically promotes expression of the polynucleotide encoding RECAP may be therapeutically useful.

30 At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based  
35 on chemical and/or structural properties of the target polynucleotide; and selection from a library of

chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding RECAP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding RECAP are assayed by any method

5 commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding RECAP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test

10 compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A

15 particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for

20 use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

25 Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition which generally comprises an active ingredient formulated with a pharmaceutically

30 acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such pharmaceutical compositions may consist of RECAP, antibodies to RECAP, and mimetics, agonists, antagonists, or inhibitors of RECAP.

35 The pharmaceutical compositions utilized in this invention may be administered by any number

of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Pharmaceutical compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of pharmaceutical compositions may be prepared for direct intracellular delivery of macromolecules comprising RECAP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, RECAP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example RECAP or fragments thereof, antibodies of RECAP, and agonists, antagonists or inhibitors of RECAP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range



of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu$ g to 100,000  $\mu$ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

#### DIAGNOSTICS

In another embodiment, antibodies which specifically bind RECAP may be used for the diagnosis of disorders characterized by expression of RECAP, or in assays to monitor patients being treated with RECAP or agonists, antagonists, or inhibitors of RECAP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for RECAP include methods which utilize the antibody and a label to detect RECAP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring RECAP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of RECAP expression. Normal or standard values for RECAP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibody to RECAP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of RECAP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding RECAP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and

quantify gene expression in biopsied tissues in which expression of RECAP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of RECAP, and to monitor regulation of RECAP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding RECAP or closely related molecules may be used to identify nucleic acid sequences which encode RECAP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding RECAP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the RECAP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:23-44 or from genomic sequences including promoters, enhancers, and introns of the RECAP gene.

Means for producing specific hybridization probes for DNAs encoding RECAP include the cloning of polynucleotide sequences encoding RECAP or RECAP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding RECAP may be used for the diagnosis of disorders associated with expression of RECAP. Examples of such disorders include, but are not limited to, a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, Down's syndrome, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia

gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; an immunological disorder, including autoimmune/inflammatory disorders, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, hematopoietic cancers, including lymphoma, leukemia, and myeloma, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding RECAP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered RECAP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding RECAP may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide

sequences encoding RECAP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample  
5 then the presence of altered levels of nucleotide sequences encoding RECAP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of RECAP,  
10 a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding RECAP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard  
15 values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to  
20 approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical  
25 symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding RECAP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically,  
30 or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding RECAP, or a fragment of a polynucleotide complementary to the polynucleotide encoding RECAP, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

35 In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences

encoding RECAP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers  
5 derived from the polynucleotide sequences encoding RECAP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently  
10 labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and  
15 automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of RECAP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from  
20 standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

25 In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described in Seilhamer, J.J. et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484, incorporated herein by reference. The microarray may also be used to  
30 identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective  
35 treatment regimen for that patient. For example, therapeutic agents which are highly effective and

display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, antibodies specific for RECAP, or RECAP or fragments thereof may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

5 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number  
10 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would  
15 provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention  
20 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-  
25 113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as  
30 the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02  
35 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at

<http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated  
5 biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

10 Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a  
15 given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the  
20 gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density  
25 related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

30 A proteomic profile may also be generated using antibodies specific for RECAP to quantify the levels of RECAP expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoz, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of  
35 methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-

reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding RECAP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial



chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, e.g., Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding RECAP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, RECAP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between RECAP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with RECAP, or fragments thereof, and washed. Bound RECAP is then detected by methods well known in the art. Purified RECAP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively,

non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding RECAP specifically compete with a test compound for binding RECAP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more  
5 antigenic determinants with RECAP.

In additional embodiments, the nucleotide sequences which encode RECAP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

10 Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following preferred specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, in  
15 particular U.S. Ser. No. 60/145,232, U.S. Ser. No. 60/158,578, and U.S. Ser. No. 60/165,192, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

20 RNA was purchased from Clontech or isolated from tissues described in Table 4. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or  
25 sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A<sup>+</sup>) RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was  
30 isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the  
35 recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units

5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSORT1 plasmid (Life Technologies), pcDNA2.1 plasmid (Invitrogen, Carlsbad CA), or pINCY plasmid (Incyte Genomics, Palo Alto CA). Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (PE Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (PE Biosystems) in conjunction with standard ABI protocols and base calling

software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example V.

The polynucleotide sequences derived from cDNA sequencing were assembled and analyzed using a combination of software programs which utilize algorithms well known to those skilled in the art. Table 5 summarizes the tools, programs, and algorithms used and provides applicable descriptions, references, and threshold parameters. The first column of Table 5 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score, the greater the homology between two sequences). Sequences were analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments were generated using the default parameters specified by the clustal algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

The polynucleotide sequences were validated by removing vector, linker, and polyA sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The sequences were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and PFAM to acquire annotation using programs based on BLAST, FASTA, and BLIMPS. The sequences were assembled into full length polynucleotide sequences using programs based on Phred, Phrap, and Consed, and were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length amino acid sequences, and these full length sequences were subsequently analyzed by querying against databases such as the GenBank databases (described above), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and Hidden Markov Model (HMM)-based protein family databases such as PFAM. HMM is a probabilistic approach which analyzes consensus primary structures of gene families. (See, e.g., Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.)

The programs described above for the assembly and analysis of full length polynucleotide and amino acid sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:23-44. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies were described in The Invention section above.

#### 35 IV. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel, 1995, supra, ch. 4 and 16.)

5 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$10 \quad \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by 15 (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score 20 represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

25 The results of northern analyses are reported as a percentage distribution of libraries in which the transcript encoding RECAP occurred. Analysis involved the categorization of cDNA libraries by organ/tissue and disease. The organ/tissue categories included cardiovascular, dermatologic, developmental, endocrine, gastrointestinal, hematopoietic/immune, musculoskeletal, nervous, reproductive, and urologic. The disease/condition categories included cancer, inflammation, trauma, cell 30 proliferation, neurological, and pooled. For each category, the number of libraries expressing the sequence of interest was counted and divided by the total number of libraries across all categories. Percentage values of tissue-specific and disease- or condition-specific expression are reported in Table 3.

#### V. Chromosomal Mapping of RECAP Encoding Polynucleotides

The cDNA sequences which were used to assemble SEQ ID NO:23-44 were compared with 35 sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other

implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:23-44 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 5). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Genetic map locations are reported as ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. SEQ ID NO:24 maps to chromosome 1 within the interval from 12.8 to 22.9 centiMorgans. SEQ ID NO:36 maps to chromosome 1 within the interval from 74.8 to 78.3 centiMorgans.

#### **VI. Extension of RECAP Encoding Polynucleotides**

The full length nucleic acid sequences of SEQ ID NO:23-44 were produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer, to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and  $\beta$ -mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2:

94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose mini-gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (PE Biosystems).

In like manner, the polynucleotide sequences of SEQ ID NO:23-44 are used to obtain 5' regulatory sequences using the procedure above, along with oligonucleotides designed for such extension, and an appropriate genomic library.

## VII. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:23-44 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National

Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### VIII. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Skena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Skena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### 35 Tissue or Cell Sample Preparation



Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65 °C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60 °C. The arrays are washed for 10 min at 45 °C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45 °C in a second wash buffer (0.1X SSC), and dried.

### 10 Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

25 The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a

linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

5 A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### IX. Complementary Polynucleotides

10 Sequences complementary to the RECAP-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring RECAP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of RECAP. To inhibit transcription, a  
15 complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the RECAP-encoding transcript.

#### X. Expression of RECAP

Expression and purification of RECAP is achieved using bacterial or virus-based expression  
20 systems. For expression of RECAP in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic  
25 resistant bacteria express RECAP upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of RECAP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding RECAP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid  
30 intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

35 In most expression systems, RECAP is synthesized as a fusion protein with, e.g., glutathione S-

transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from RECAP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16).

10 Purified RECAP obtained by these methods can be used directly in the assays shown in Examples XI and XV.

#### XI. Demonstration of RECAP Activity

Receptor activity of RECAP is determined in a ligand-binding assay using candidate ligand molecules in the presence of <sup>125</sup>I-labeled RECAP. RECAP is labeled with <sup>125</sup>I Bolton-Hunter reagent.

15 (See, e.g., Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate ligand molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled RECAP, washed, and any wells with labeled RECAP complex are assayed. Data obtained using different concentrations of RECAP are used to calculate values for the number, affinity, and association of RECAP with the ligand molecules. The level of binding measured is proportional to the level of RECAP activity.

20 In the alternative, activity of RECAP may be measured using an assay based upon the property of some GPCRs to support the in vitro proliferation of fibroblasts and tumor cells under serum-free conditions (Chiquet-Ehrismann, R, et al. (1986) Cell 47:131-139). Wells in 96 well cluster plates (Falcon, Fisher Scientific, Santa Clara CA) are coated with RECAP by incubation with solutions at 50-100 µg/ml for 15 min at ambient temperature. The coating solution is aspirated, and the wells

25 washed with Dulbecco's medium before cells are plated. Rat fibroblast cultures or rat mammary tumor cells are prepared as described and plated at a density of 10<sup>4</sup>-10<sup>5</sup> cells/ml in Dulbecco's medium supplemented with 10% fetal calf serum (FCS).

After three days the media are removed, and the cells washed three times with phosphate-buffered saline (PBS) before the addition of serum-free Dulbecco's medium containing 0.25 mg/ml

30 bovine serum albumin (BSA, Fraction V, Sigma Chemical, St. Louis, MO). After 2 days the medium is aspirated, and 100 µl of [<sup>3</sup>H]thymidine (NEN) at 2 µCi/ml in fresh Dulbecco's medium containing 0.25 mg/ml BSA added. Parallel plates are fixed and stained to determine cell numbers. After 16 hr, the medium is aspirated, the cell layer washed with PBS, and the 10% trichloroacetic acid-precipitable counts in the cell layer determined by liquid scintillation counting of radioisotope (normalized to

35 relative cell numbers; Chiquet-Ehrismann, R. et al. (1986) supra). The rates of cell proliferation and

[<sup>3</sup>H]thymidine uptake are proportional to the activity of RECAP in the sample.

In the alternative, the assay for RECAP activity is based upon the property of CD97/Emr1 GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P., et al. (1998) *J. Biol. Chem.*, 273:4990-4996). A plasmid encoding full length RECAP is transfected into a mammalian cell line (e.g., COS-7 or Chinese hamster ovary (CHO-K1) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium containing 2% FCS for 48 hours, the culture medium is discarded, then the attached cells are gently washed with PBS. The cells are then incubated in culture medium with 10% FCS or 2% FCS for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid.

10 The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art. Changes in the levels of cAMP in the lysate from 10% FCS-treated cells compared with those in 2% FCS-treated cells are proportional to the activity of RECAP present in the transfected cells.

In another alternative, an assay for RECAP activity is based on a prototypical assay for ligand/receptor-mediated modulation of cell proliferation. This assay measures the rate of DNA synthesis in Swiss mouse 3T3 cells. A plasmid containing polynucleotides encoding RECAP is added to quiescent 3T3 cultured cells using transfection methods well known in the art. The transiently transfected cells are then incubated in the presence of [<sup>3</sup>H]thymidine, a radioactive DNA precursor molecule. Varying amounts of RECAP ligand are then added to the cultured cells. Incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA is measured over an appropriate time interval using a radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold RECAP ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of RECAP producing a 50% response level, where 100% represents maximal incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA (McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY, p. 73.)

20 radioisotope counter, and the amount incorporated is directly proportional to the amount of newly synthesized DNA. A linear dose-response curve over at least a hundred-fold RECAP ligand concentration range is indicative of receptor activity. One unit of activity per milliliter is defined as the concentration of RECAP producing a 50% response level, where 100% represents maximal incorporation of [<sup>3</sup>H]thymidine into acid-precipitable DNA (McKay, I. and Leigh, I., eds. (1993) Growth Factors: A Practical Approach, Oxford University Press, New York, NY, p. 73.)

In the alternative, the assay for RECAP activity is based upon the ability of GPCR family proteins to modulate G protein-activated second messenger signal transduction pathways (e.g., cAMP; Gaudin, P. et al. (1998) *J. Biol. Chem.* 273:4990-4996). A plasmid encoding full length RECAP is transfected into a mammalian cell line (e.g., Chinese hamster ovary (CHO) or human embryonic kidney (HEK-293) cell lines) using methods well-known in the art. Transfected cells are grown in 12-well trays in culture medium for 48 hours, then the culture medium is discarded, and the attached cells are gently washed with PBS. The cells are then incubated in culture medium with or without ligand for 30 minutes, then the medium is removed and cells lysed by treatment with 1 M perchloric acid. The cAMP levels in the lysate are measured by radioimmunoassay using methods well-known in the art.

35 Changes in the levels of cAMP in the lysate from cells exposed to ligand compared to those without

ligand are proportional to the amount of RECAP present in the transfected cells.

## **XII. Functional Assays**

RECAP function is assessed by expressing the sequences encoding RECAP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression  
5 vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include pCMV SPORT (Life Technologies) and pCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are  
10 co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other  
15 cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins  
20 as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of RECAP on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding RECAP and either CD64 or CD64-GFP. CD64 and CD64-  
25 GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding RECAP and other genes of interest can be analyzed by northern analysis or  
30 microarray techniques.

## **XIII. Production of RECAP Specific Antibodies**

RECAP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

35 Alternatively, the RECAP amino acid sequence is analyzed using LASERGENE software

(DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

5 Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (PE Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-RECAP  
10 activity by, for example, binding the peptide or RECAP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XIV. Purification of Naturally Occurring RECAP Using Specific Antibodies**

Naturally occurring or recombinant RECAP is substantially purified by immunoaffinity chromatography using antibodies specific for RECAP. An immunoaffinity column is constructed by  
15 covalently coupling anti-RECAP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing RECAP are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of RECAP (e.g., high ionic strength buffers in the  
20 presence of detergent). The column is eluted under conditions that disrupt antibody/RECAP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and RECAP is collected.

#### **XV. Identification of Molecules Which Interact with RECAP**

RECAP, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent.  
25 (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled RECAP, washed, and any wells with labeled RECAP complex are assayed. Data obtained using different concentrations of RECAP are used to calculate values for the number, affinity, and association of RECAP with the candidate molecules.

30 Alternatively, molecules interacting with RECAP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989, Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

RECAP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions  
35 between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent

No. 6,057,101).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention.

- 5 Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.



Table 1

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
1	23	209171	SPLNNOT02	156553R6 (THP1PLB02), 209171F1 (SPLNNOT02), 209171H1 (SPLNNOT02), 341273R6 (NEUTFMT01), 607227X11 (BRSTTUT01), 921863X30R1 (RATRN0202)
2	24	945430	RATRN0202	465647X19C1 (LATRN0201), 945430H1 (RATRN0202), 2886970F6 (SINJN0202), SAIA01782F1, SAIA03901F1, SAIA00918F1, SAIA03865F1
3	25	1305513	PLACN0202	1305513H1 (PLACN0202), SXZA00319V1, SXZA00758V1, SXZA00011V1, SXZA00705V1, SXZA00520V1, SXZA00707V1, SXZA00525V1
4	26	1876283	LEUKN0202	1520713F1 (BLADTUT04), 1815520T6 (PROSN020), 1876283H1 (LEUKN0202), 1876283X310D1 (LEUKN0202)
5	27	2470285	THP1N0T03	2470285H1 (THP1N0T03), 2470285X26C1 (THP1N0T03), 2470285X304U1 (THP1N0T03), 2470285X313B1 (THP1N0T03), 2470285X31C1 (THP1N0T03), 2470285X41C1 (THP1N0T03), 2470285X44C1 (THP1N0T03), 2470285X46C1 (THP1N0T03), 4874704H1 (COLDN0T01), g3960000, g965238, g2063924
6	28	2925789	SININ0T04	722886R1 (SYNCOAT01), 955207R7 (KIDN0T05), 1336911T1 (COLN0T13), 1854413F6 (HNT3AZT01), 2196369F6 (SPLNFET02), 2275751R6 (PROSN0N01), 2925789F6 (SININ0T04), 2925789H1 (SININ0T04), 3538492H1 (SEMVN0T04)
7	29	3099990	STOMFET02	1824381H1 (GBLATUT01), 2553230H1 (THYMN0T03), 3099990H1 (STOMFET02), 3268969H1 (BRAIN0T20), g3155644, g1491543
8	30	103561	BMARN0T02	103561H1 (BMARN0T02), SBBA01615F1, g3836278
9	31	288709	EOSIHET02	288709F1 (EOSIHET02), 288709H1 (EOSIHET02), 3393757X301D2 (LUNGNOT28), 3395207F6 (LUNGNOT28), 4413060F6 (MONOTXT01), 4413060T6 (MONOTXT01), SZAH04055F1
10	32	959893	BRSTTUT03	959893H1 (BRSTTUT03), 959893R6 (BRSTTUT03)
11	33	1414179	BRAIN0T12	1272762F1 (TESTTUT02), 2121559T6 (BRSTN0T07), 3248471H1 (SEMVN0T03), 4324516H1 (TLYMUNT01)
12	34	2197211	SPLNFET02	2197211F6 (SPLNFET02), 2197211H1 (SPLNFET02)
13	35	2263653	UTRSN0T02	140819X2 (TLYMN0R01), 1550714T6 (PROSN0T06), 1843270R6 (COLN0T08), 1906033F6 (OVARN0T07), 2110044R6 (BRAITUT03), 2263653H1 (UTRSN0T02), 4596808H1 (COLSTUT01), 4891416H1 (PROSTMT05), 5063684F6 (ARTFTDT01)

Table 1 (cont.)

Polypeptide SEQ ID NO:	Nucleotide SEQ ID NO:	Clone ID	Library	Fragments
14	36	2504590	CONUTUT01	1428502T1 (SINTBST01), 2504590H1 (CONUTUT01), SAJA00914R1, SAJA00733R1, SAJA00921R1
15	37	2529619	GBLANOT02	1504422X17C1 (BRAITUT07), 1506560X26C1 (BRAITUT07), 1516974F6 (PANCUT01), 2529619H1 (GBLANOT02)
16	38	5467661	LNODNOT11	2310518R6 (NGANNOT01), 2640268F6 (LUNGUTUT08), 2893053F6 (LUNGFET04), 3183381H1 (OVARNOT11), 3343709F6 (SPLNNOT09), 5049433T6 (BRSTNOT33), SBAA04161F2
17	39	229740	PANCNOT01	032924H1 (THP1NOB01), 229740H1 (PANCNOT01), 229740R1 (PANCNOT01), 881634R1 (THYRNOT02), 2072921F6 (ISLTNOT01), 2072921T6 (ISLTNOT01), 2614287H1 (GBLANOT01), 3362830H1 (PROSBPT02), 3409621H1 (PROSTUS08), g4249643
18	40	1317467	BLADTUT02	035646H1 (HUVENOB01), 412620R1 (BRSTNOT01), 1317467F6 (BLADTUT02), 1317467H1 (BLADTUT02), 2023272F6 (CONNNOT01), 2023272T6 (CONNNOT01), 2457956H1 (ENDANOT01), 4459319H1 (HEAADIT01), 4834580H1 (BRAWNOT01), 5097717H1 (EPIMNON05), 5293601H2 (COLENOT01)
19	41	2279267	PROSNON01	2279267H1 (PROSNON01), 3001127F6 (TLYMNOT06), 3425035H1 (BRSTNOR01)
20	42	2436258	BRAVUNT02	633426H1 (NEUTGMT01), 1984786R6 (LUNGAST01), 2436258H1 (BRAVUNT02), 4109419F6 (PROSBPT07), 4594456H1 (PROSTUT18), g1349289
21	43	2681738	SINIUCT01	775882R1 (COLNNOT05), 1752341F6 (LIVRTUT01), 2520558F6 (BRAITUT21), 2681738F6 (SINIUCT01), 2681738H1 (SINIUCT01), 3389931F6 (LUNGUTUT17), 4379601H1 (LUNGNOT37)
22	44	2859482	SININOT03	161339H1 (ADENINB01), 573392H1 (BRAVUNT01), 1002066H1 (BRSTNOT03), 1992904H1 (CORPNOT02), 2209522H1 (SINTFET03), 2257029R6 (OVARUT01), 2620749R6 (KERANOT02), 2859482H1 (SININOT03), 2859867F6 (SININOT03), 3000455H1 (TLYMNOT06), 3106558H1 (BRSTTUT15), 3970970H1 (PROSTUT10), 5687790H1 (BRAIUNT01), g4582148

Table 2

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
1	411	S75 T101 S129 S130 S143 T207 T235 T245 S294 S319 S329 T362 S376 S35 S72 T118 S119 T227 S289			Retinoid X receptor interacting protein [Homo sapiens] g6523831 Hillier L.D. et al. (1996) Genome Res 6:807-28.	BLAST-GenBank MOTIFS
2	579	T16 T59 T60 S163 T525 S69 T120 T130 S135 T209 S248 T277 T311 S474 T503 Y513	N81 N416 N501 N543	Signal peptide: M1-V25	Human retinol binding protein receptor R44617	SPScan BLAST-GENESEQ BLAST-DOMO MOTIFS
3	370	T134 S284 S342 S80 T93 T130 S178 T266	N3 N83 N182 N227 N264	G protein-coupled receptor signature: I30-S351 Rhodopsin-like GPCR superfamily: L22-V46, P55-F76, G101-I238, T137-L158, I283-W307, L321-R347 Transmembrane domains: L24-I41, C105-A122, D183-L203	G protein-coupled receptor [Mus musculus] g2739105	BLAST-GenBank BLAST-DOMO BLAST-PRODROM HMNER HMNER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS MOTIFS
4	267	S65 T210 S217	N208	Signal Peptide: M1-G56 Transmembrane domains: L45-Y61, L179-L196 Tumor Necrosis Factor receptor family cysteine-rich signature: C99-G135	Putative ankyrin repeat-containing protein [Mortierella alpina] g5921507	SPScan HMNER HMNER-PFAM MOTIFS BLAST-GENBANK

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
5	951	T820 S143 S164 S191 S249 T416 S421 S488 S508 T595 S646 T856 S44 S133 S390 S572 S646 T820 S871 Y352	N68 N199 N294 N314 N505 N854	G protein-coupled receptor signature: P531-L815 Transmembrane domains: L543-I560, L704-I724, V749-I775	G-protein-coupled receptor [Homo sapiens] g7739737	BLAST-GenBank BLAST-DOMO HMNER HMNER-PFAM BLIMPS-PRINTS MOTIFS
6	413	T236 S240 S376 T180 S315 Y252	N63 N234	Transmembrane domains: W22-D41, T145-L170, T205-I226 Tumor Necrosis Factor receptor family cysteine-rich signature: C101-C136		HMNER HMNER-PFAM MOTIFS
7	144	S44 S82 T19 S94 S111 T131		Calcitonin receptor signature: R110-A124	Receptor like protein (fragment) [Arabidopsis thaliana] g3046693	BLAST-GenBank BLIMPS-PRINTS MOTIFS
8	174	S95 S30 S86 S13 S70	N48 N170	Signal peptide: M1-Q34 Sushi domains: C35-C91; C96-C153 Complement factor H repeat: Q34-S95; K88-D154 Complement pathway membrane protein domain: M1-S95	Complement receptor 1 [Homo sapiens] g563324	BLAST-GenBank MOTIFS SPSCAN HMNER HMNER-PFAM BLIMPS-PFAM BLAST-PRODOM BLAST-DOMO

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
9	449	S332 S71 S416 S418 S436 S87 T244 S426 Y423	N73 N77 N183 N247 N252	Transmembrane domains: M159-L177; W262-T289 M302-I326; Y378-L398 G-protein coupled receptors family 2 signature: C216-L241; G268-R292 W303-S332; V369-E412 Secretin-like GPCR superfamily signature: V155-K179; I218-L241 K261-L286; W303-K328 A377-L398 CD97/EMR1 receptors domain: S63-K434 CD97 GPCR domain: M1-V146	EGF-like module EMR2 [Homo sapiens] g6650689	BLAST-GenBank MOTIFS HMMER BLIMPS-BLOCKS PROFILES-SCAN BLIMPS-PRINTS BLAST-PRODOM BLAST-DOMO
10	126	S21 T89	N44	Signal peptide: M1-S21 Immunoglobulin domain: G36-L112; E25-S93	TCRAV6S1 (T-cell receptor alpha chain) [Homo sapiens] g2358027	BLAST-GenBank MOTIFS SPSCAN HMMER HMMER-PFAM BLAST-DOMO
11	273	S25 S41 S54 S94 S66 S77 S93 S9 S17 S46 S90 T130 S268		Opioid receptor signature: R40-R52	Thyrotropin G protein-coupled receptor N-terminal sequence [Homo sapiens] Geneseq ID W03626	BLAST-Geneseq MOTIFS BLIMPS-PRINTS

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
12	140	S92 S20 S73 T88 Y107	N43	Signal peptide: M1-G21 Immunoglobulin domain: G37-V111 T-cell receptor alpha chain signature: C7-P131 T-cell surface antigen domain: F9-P138	T-cell receptor alpha chain [Macaca mulatta] g555729	BLAST-GenBank MOTIFS SPSCAN HMMER HMMER-PFAM BLAST-PRODOM BLAST-DOMO
13	479	S44 T90 S160 T252 T258 S309 S422 S147 S313	N34 N387	Transmembrane domains: V169-V187; L225-G246 L454-F472 Delta opioid receptor signature: A328-L340; P404-S416		MOTIFS HMMER BLIMPS-PRINTS
14	99	S91			Alpha 1C adrenergic receptor isoform 2 [Homo sapiens] g927209	BLAST-GenBank MOTIFS
15	349	T307 T140 S338	N8 N45	Transmembrane domain: I26-G44; F203-V219 7 TM receptor domain: G44-Y293 G-protein coupled receptor signature: K93-P132; N285-R301 P24-R301 Olfactory receptor signature: M62-Q83; F180-D194 F241-G256; L277-L288 G155-R301	Similar to mouse olfactory receptor [Homo sapiens] g4159884	BLAST-GenBank MOTIFS HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS PROFILES SCAN BLAST-PRODOM BLAST-DOMO

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
16	373	T3 T111 S179 T336 T363 T40 S67 S147 S224 S293 S365	N11 N23 N361	Transmembrane domains: P78-M102; I120-G140 F193-L211; F228-F251	Seven transmembrane domain orphan receptor 3 [Homo sapiens] g6729336	BLAST-GenBank MOTIFS HMMER
17	353	S273 T146 S163 T188 S281 T309 S327 T18 T30 S54 T188 S287 S306 Y316	N68 N74 N79 N136 N144	WH1 domain: E13-K117 Coiled coil repeat: E103-L332 Leucine zipper: L325-L346	glutamate receptor associated protein homer-2b [Homo sapiens] g3834619 (Tu, J.C. et al. (1998) Neuron 21:717-726.)	BLAST-GenBank HMMER-PFAM BLAST-PRODOM MOTIFS
18	441	S104 T167 S203 T266 S372 S382 S402 S427 S99 S104 S148 S155 S202 S223 S278 S365 Y286	N62 N165	Signal peptide: M1-S43 P2Y6 purinoreceptor: E197-C213 SP1a and ryanodine receptor (SPRY) domain: E369-S382	predicted G-protein coupled receptor [C. elegans] g3876583	BLAST-GenBank SPSCAN BLIMPS-PRINTS BLIMPS-PFAM MOTIFS

Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
19	310	S7 T136 S290 T299	N4 N41	Transmembrane domain: I22-G40 7 transmembrane receptor domain: G40-C289 GPCR domain: K89-P128, N281-K297 Olfactory receptor signature: M58-R79, F176-D190, F237-G252, S290-L304, L165-L244 Melanocortin receptor family: L50-L62, I125-T136 Vasopressin receptor signature: L54-L65	odorant receptor [Mus musculus] g293754 (Ressler, K.J. et al. (1993) 73:597-609.)	BLAST-GenBank HMMER HMMER-PFAM BLIMPS-BLOCKS BLIMPS-PRINTS BLAST-DOMO BLAST-PRODOM MOTIFS
20	438	T160 T246 T322 S331 S375 T424 S116 T246 T353 T374 Y228	N282		Sand (plasminogen related growth factor receptor) (Fugu rubripes) g3928166	BLAST-GenBank MOTIFS



Table 2 (cont.)

Seq ID NO:	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Motifs, and Domains	Homologous Sequences	Analytical Methods and Databases
21	357	T4 S301 S59	N158	Transmembrane domains: M64-A84, V178-F197, L131-E151, Y214-P234, F99-V117 Glutamate receptor: G102-V123, R208-T229 Transmembrane 4 family: T96-L119, N174-S202, I87-L113 Muscarinic M2 receptor: S336-V352		HMMER BLIMPS-PRINTS MOTIFS
22	1069	T448 T488 T489 S931 S42 S86 S163 T203 T337 T399 T409 S434 S447 T470 S479 S481 S508 T540 S600 T623 S639 T766 S767 T774 T823 S987 T996 S270 T337 T399 S444 S481 S493 T733 T766 S810 T823 T865 S945 S987 T1002 S1056 Y306 Y379 Y472 Y821	N40 N54 N190 N466 N611 N930 N1051	TBC GTPase activation domain: V563-T774 rabGAP domain: I606-P615, Y647-S652 (P<2.2e-3) Phosphotyrosine interaction domain: F147-K465 Membrane protein family: W541-I756 Leucine zipper: L538-L559 P loop (ATP/GTP binding site A): G371-S378	predicted rabGAP domain protein [C. elegans] g1109865 (Siderovski, D.P. et al. (1999) 34:215- 251)	BLAST-GenBank HMMER-PFAM BLIMPS-PFAM BLAST-PRODOM BLAST-DOMO MOTIFS

Table 3

Nucleotide Seq ID NO:	Selected Fragment(s)	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
23	607-663	Hematopoietic/Immune (0.333) Reproductive (0.200) Developmental (0.100) Musculoskeletal (0.100)	Inflammation (0.433) Cancer (0.333) Cell Proliferation (0.233)	PBLUESCRIPT
24	890-934 1277-1321	Gastrointestinal (0.333) Cardiovascular (0.250) Nervous (0.167) Reproductive (0.167)	Inflammation (0.500) Cancer (0.250)	PSPORT1
25	748-792 1582-1626	Developmental (0.250) Endocrine (0.250) Nervous (0.250) Reproductive (0.250)	Cell Proliferation (0.500) Cancer (0.250) Inflammation (0.250)	pINCY
26	248-292	Reproductive (0.238) Hematopoietic/Immune (0.190) Gastrointestinal (0.175)	Cancer (0.508) Inflammation (0.301) Cell Proliferation (0.238)	pINCY
27	1474-1518	Reproductive (0.393) Nervous (0.179) Gastrointestinal (0.179)	Cancer (0.643) Inflammation (0.179) Cell Proliferation (0.107)	pINCY
28	1595-1645	Reproductive (0.235) Gastrointestinal (0.176) Hematopoietic/Immune (0.147)	Cancer (0.485) Inflammation (0.353) Cell Proliferation (0.147)	pINCY
29	31-75 535-579	Developmental (0.400) Nervous (0.200) Gastrointestinal (0.200) Hematopoietic/Immune (0.200)	Cell Proliferation (0.400) Cancer (0.200) Neurological (0.200)	pINCY
30	15-59	Reproductive (0.250) Hematopoietic/Immune (0.250) Gastrointestinal (0.167) Nervous (0.167)	Cancer (0.500) Inflammation/Trauma (0.333) Cell proliferation (0.083)	PBLUESCRIPT
31	372-416 1530-1574	Hematopoietic/Immune (0.500) Cardiovascular (0.333) Gastrointestinal (0.167)	Inflammation/Trauma (0.500) Cancer (0.167)	PBLUESCRIPT
32	386-430	Cardiovascular (0.286) Gastrointestinal (0.286) Hematopoietic/Immune (0.286)	Cancer (0.571) Inflammation/Trauma (0.143)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
33	703-747	Reproductive (0.260) Gastrointestinal (0.193) Nervous (0.127)	Cancer (0.427) Inflammation/Trauma (0.306) Cell proliferation (0.173)	pINCY
34	398-442	Reproductive (0.667) Developmental (0.333)	Cancer (0.667) Cell proliferation (0.333)	pINCY
35	542-586 974-1018	Reproductive (0.294) Nervous (0.157) Gastrointestinal (0.137)	Cancer (0.510) Inflammation/Trauma (0.294) Cell proliferation (0.255)	PSPORT1
36	279-323	Reproductive (0.333) Gastrointestinal (0.167) Hematopoietic/Immune (0.167) Urologic (0.167)	Cancer (0.500) Inflammation/Trauma (0.500)	pINCY
37	919-963	Reproductive (0.467) Cardiovascular (0.133) Gastrointestinal (0.100) Nervous (0.100)	Cancer (0.600) Inflammation/Trauma (0.274) Cell proliferation (0.133)	pINCY
38	1313-1357	Reproductive (0.233) Hematopoietic/Immune (0.150) Cardiovascular (0.117) Developmental (0.117)	Inflammation/Trauma (0.366) Cancer (0.350) Cell proliferation (0.300)	pINCY
39	1-45	Reproductive (0.455) Gastrointestinal (0.227)	Cancer (0.318) Inflammation (0.273) Cell proliferation (0.182) Trauma (0.182)	PBLUESCRIPT
40	127-171 481-525 757-801	Reproductive (0.320) Nervous (0.240) Gastrointestinal (0.200)	Cancer (0.360) Inflammation (0.240) Trauma (0.160)	pINCY
41	928-972	Reproductive (0.333) Cardiovascular (0.167) Nervous (0.167) Gastrointestinal (0.167) Hematopoietic/Immune (0.167)	Cancer (0.500) Inflammation (0.333) Trauma (0.167)	PSPORT1

Table 3 (cont.)

Nucleotide SEQ ID NO:	Selected Fragments	Tissue Expression (Fraction of Total)	Disease or Condition (Fraction of Total)	Vector
42	21-65	Reproductive (0.455) Hematopoietic/Immune (0.182) Nervous (0.182)	Cancer (0.545) Inflammation (0.182)	PSPORT1
43	1-45	Gastrointestinal (0.275) Cardiovascular (0.225) Reproductive (0.175)	Cancer (0.475) Inflammation (0.325) Cell proliferation (0.125)	pINCY
44	202-246	Reproductive (0.419) Nervous (0.129) Hematopoietic/Immune (0.097)	Cancer (0.516) Cell proliferation (0.161) Inflammation (0.161)	pINCY

Table 4

Nucleotide SEQ ID NO:	Library	Library Description
23	SPLNNOT02	Library was constructed using RNA isolated from the spleen of a 29-year-old Caucasian male, who died from head trauma. Serologies were positive for cytomegalovirus (CMV).
24	RATRNOT02	Library was constructed using RNA isolated from the right atrium tissue of a 39-year-old Caucasian male, who died from a gunshot wound.
25	PLACNOT02	Library was constructed using RNA isolated from the placental tissue of a Hispanic female fetus, who was prematurely delivered at 21 weeks' gestation. Serologies of the mother's blood were positive for CMV (cytomegalovirus).
26	LEUKNOT02	Library was constructed using RNA isolated from white blood cells of a 45-year-old female with blood type O+. The donor tested positive for cytomegalovirus (CMV).
27	THPINOT03	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 (ATCC TIB 202) is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia.
28	SININOT04	Library was constructed using RNA isolated from diseased ileum tissue obtained from a 26-year-old Caucasian male during a partial colectomy, permanent colostomy, and an incidental appendectomy. Pathology indicated moderately to severely active Crohn's disease. Family history included enteritis of the small intestine.
29	STOMFET02	Library was constructed using RNA isolated from stomach tissue removed from a Hispanic male fetus, who died at 18 weeks' gestation.
30	BMARNOT02	This library was constructed using RNA isolated from the bone marrow of 24 male and female Caucasian donors, 16 to 70 years old. (RNA came from Clontech.)
31	EOSIHET02	This library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hypereosinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
32	BRSTTUT03	This library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast. Patient history included skin cancer, rheumatic heart disease, osteoarthritis, and tuberculosis. Family history included cerebrovascular disease, coronary artery aneurysm, breast cancer, prostate cancer, atherosclerotic coronary artery disease, and type I diabetes.

Table 4 (cont.)

Nucleotide SEQ ID NO.:	Library	Library Description
33	BRAINOT12	This library was constructed using RNA isolated from brain tissue removed from the right frontal lobe of a 5-year-old Caucasian male during a hemispherectomy. Pathology indicated extensive polymicrogyria and mild to moderate gliosis (predominantly subpial and subcortical), which are consistent with chronic seizure disorder. Family history included a cervical neoplasm.
34	SPLNFET02	This library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.
35	UTRSNOT02	This library was constructed using RNA isolated from uterine tissue removed from a 34-year-old Caucasian female during a vaginal hysterectomy. Patient history included mitral valve disorder. Family history included stomach cancer, congenital heart anomaly, irritable bowel syndrome, ulcerative colitis, colon cancer, cerebrovascular disease, type II diabetes, and depression.
36	CONUTUT01	This library was constructed using RNA isolated from sigmoid mesentery tumor tissue obtained from a 61-year-old female during a total abdominal hysterectomy and bilateral salpingo-oophorectomy with regional lymph node excision. Pathology indicated a metastatic grade 4 malignant mixed mullerian tumor present in the sigmoid mesentery at two sites.
37	GBLANOT02	This library was constructed using RNA isolated from diseased gallbladder tissue removed from a 21-year-old Caucasian male during a cholecystectomy. Pathology indicated moderate chronic cholecystitis, cholelithiasis with 1 mixed stone, and acute serositis. Family history included benign hypertension, breast cancer, colon cancer, and type II diabetes.
38	LNODNOT11	This library was constructed using RNA isolated from lymph node tissue removed from a 16-month-old Caucasian male who died from head trauma. Patient history included bronchitis.
39	PANCNOT01	This library was constructed using RNA isolated from the pancreatic tissue of a 29-year-old Caucasian male who died from head trauma.
40	BLADTUT02	This library was constructed using RNA isolated from bladder tumor tissue removed from an 80-year-old Caucasian female. Pathology indicated invasive transitional cell carcinoma. Family history included acute renal failure, osteoarthritis, and atherosclerosis.
41	PROSNON01	This normalized prostate library was constructed from 4.4 million independent clones from a prostate library. Starting RNA was made from prostate tissue removed from a 28-year-old Caucasian male who died from a self-inflicted gunshot wound. The normalization and hybridization conditions were adapted from Soares, M.B. et al. (1994) Proc. Natl. Acad. Sci. USA 91:9228-9232, using a longer (19 hour) reannealing hybridization period.

Table 4 (cont)

Nucleotide SEQ ID NO:	Library	Library Description
42	BRAVUNT02	This library was constructed using RNA isolated from separate populations of unstimulated astrocytes.
43	SINIUCT01	This library was constructed using RNA isolated from ileum tissue obtained from a 42-year-old Caucasian male. Family history included cerebrovascular disease, benign hypertension, atherosclerotic coronary artery disease, and type II diabetes.
44	SININOT03	This library was constructed using RNA isolated from ileum tissue obtained from an 8-year-old Caucasian female, who died from head trauma. Serology was positive for cytomegalovirus (CMV).

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

- 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:22,
- b) a naturally occurring amino acid sequence having at least 90% sequence identity to an
- 10 amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:22,
- c) a biologically active fragment of an amino acid sequence selected from the group consisting
- 15 of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:22, and
- d) an immunogenic fragment of an amino acid sequence selected from the group consisting of
- 20 SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:22.

- 25 2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:22.

30

3. An isolated polynucleotide encoding a polypeptide of claim 1.

4. An isolated polynucleotide encoding a polypeptide of claim 2.

- 35 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:23,



SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:44.

5

6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

7. A cell transformed with a recombinant polynucleotide of claim 6.

10

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

9. A method for producing a polypeptide of claim 1, the method comprising:

- 15 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
- b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

20

11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- 25 a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:44,
- b) a naturally occurring polynucleotide sequence having at least 70% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, and SEQ ID NO:44,
- 30 c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- 35 e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide  
5 having a sequence of a polynucleotide of claim 11, the method comprising:  
a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and  
10 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide  
15 having a sequence of a polynucleotide of claim 11, the method comprising:  
a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and  
b) detecting the presence or absence of said amplified target polynucleotide or fragment  
20 thereof, and, optionally, if present, the amount thereof.

16. A pharmaceutical composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

17. A pharmaceutical composition of claim 16, wherein the polypeptide comprises an amino  
25 acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, and SEQ ID NO:22.

18. A method for treating a disease or condition associated with decreased expression of functional RECAP, comprising administering to a patient in need of such treatment the pharmaceutical composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of

claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

5           20. A pharmaceutical composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

          21. A method for treating a disease or condition associated with decreased expression of functional RECAP, comprising administering to a patient in need of such treatment a pharmaceutical  
10 composition of claim 20.

          22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- 15 b) detecting antagonist activity in the sample.

          23. A pharmaceutical composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

20           24. A method for treating a disease or condition associated with overexpression of functional RECAP, comprising administering to a patient in need of such treatment a pharmaceutical composition of claim 23.

          25. A method of screening for a compound that specifically binds to the polypeptide of claim  
25 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

30

          26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- 35 b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound,

and

- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- 10      a) exposing a sample comprising the target polynucleotide to a compound, and  
b) detecting altered expression of the target polynucleotide.

28. A method for assessing toxicity of a test compound, said method comprising:

- 15      a) treating a biological sample containing nucleic acids with the test compound;  
b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;  
20      c) quantifying the amount of hybridization complex; and  
d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

## SEQUENCE LISTING

<110> INCYTE GENOMICS, INC.  
 AU-YOUNG, Janice  
 BANDMAN, Olga  
 TANG, Y. Tom  
 YUE, Henry  
 AZIMZAI, Yalda  
 BURFORD, Neil  
 BAUGHN, Mariah R.  
 LU, Dyung Aina M.  
 HILLMAN, Jennifer L.  
 PATTERSON, Chandra  
 LAL, Preeti

<120> RECEPTORS AND ASSOCIATED PROTEINS

<130> PF-0726 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/145,232; 60/158,578; 60/165,192

<151> 1999-07-21; 1999-10-07; 1999-11-12

<160> 44

<170> PERL Program

<210> 1

<211> 411

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 209171CD1

<400> 1

Met	Ala	Gln	Arg	Gln	Leu	Leu	Asn	Lys	Lys	Gly	Phe	Gly	Glu	Pro
1				5					10					15
Val	Leu	Pro	Arg	Pro	Pro	Ser	Leu	Ile	Gln	Asn	Glu	Cys	Gly	Gln
				20					25					30
Gly	Glu	Gln	Ala	Ser	Glu	Lys	Asn	Glu	Cys	Ile	Ser	Glu	Asp	Met
				35					40					45
Gly	Asp	Glu	Asp	Lys	Glu	Glu	Arg	Gln	Glu	Ser	Arg	Ala	Ser	Asp
				50					55					60
Trp	His	Ser	Lys	Thr	Lys	Asp	Phe	Gln	Glu	Ser	Ser	Ile	Lys	Ser
				65					70					75
Leu	Lys	Glu	Lys	Leu	Leu	Leu	Glu	Glu	Glu	Pro	Thr	Thr	Ser	His
				80					85					90
Gly	Gln	Ser	Ser	Gln	Gly	Ile	Val	Glu	Glu	Thr	Ser	Glu	Glu	Gly
				95					100					105
Asn	Ser	Val	Pro	Ala	Ser	Gln	Ser	Val	Ala	Ala	Leu	Thr	Ser	Lys
				110					115					120
Arg	Ser	Leu	Val	Leu	Met	Pro	Glu	Ser	Ser	Ala	Glu	Glu	Ile	Thr
				125					130					135
Val	Cys	Pro	Glu	Thr	Gln	Leu	Ser	Ser	Ser	Glu	Thr	Phe	Asp	Leu
				140					145					150
Glu	Arg	Glu	Val	Ser	Pro	Gly	Ser	Arg	Asp	Ile	Leu	Asp	Gly	Val
				155					160					165
Arg	Ile	Ile	Met	Ala	Asp	Lys	Glu	Val	Gly	Asn	Lys	Glu	Asp	Ala
				170					175					180

Glu	Lys	Glu	Val	Ala	Ile	Ser	Thr	Phe	Ser	Ser	Ser	Asn	Gln	Val	
				185					190					195	
Ser	Cys	Pro	Leu	Cys	Asp	Gln	Cys	Phe	Pro	Pro	Thr	Lys	Ile	Glu	
				200					205					210	
Arg	His	Ala	Met	Tyr	Cys	Asn	Gly	Leu	Met	Glu	Glu	Asp	Thr	Val	
				215					220					225	
Leu	Thr	Arg	Arg	Gln	Lys	Glu	Ala	Lys	Thr	Lys	Ser	Asp	Ser	Gly	
				230					235					240	
Thr	Ala	Ala	Gln	Thr	Ser	Leu	Asp	Ile	Asp	Lys	Asn	Glu	Lys	Cys	
				245					250					255	
Tyr	Leu	Cys	Lys	Ser	Leu	Val	Pro	Phe	Arg	Glu	Tyr	Gln	Cys	His	
				260					265					270	
Val	Asp	Ser	Cys	Leu	Gln	Leu	Ala	Lys	Ala	Asp	Gln	Gly	Asp	Gly	
				275					280					285	
Pro	Glu	Gly	Ser	Gly	Arg	Ala	Cys	Ser	Thr	Val	Glu	Gly	Lys	Trp	
				290					295					300	
Gln	Gln	Arg	Leu	Lys	Asn	Pro	Lys	Glu	Lys	Gly	His	Ser	Glu	Gly	
				305					310					315	
Arg	Leu	Leu	Ser	Phe	Leu	Glu	Gln	Ser	Glu	His	Lys	Thr	Ser	Asp	
				320					325					330	
Ala	Asp	Ile	Lys	Ser	Ser	Glu	Thr	Gly	Ala	Phe	Arg	Val	Pro	Ser	
				335					340					345	
Pro	Gly	Met	Glu	Glu	Ala	Gly	Cys	Ser	Arg	Glu	Met	Gln	Ser	Ser	
				350					355					360	
Phe	Thr	Arg	Arg	Asp	Leu	Asn	Glu	Ser	Pro	Val	Lys	Ser	Phe	Val	
				365					370					375	
Ser	Ile	Ser	Glu	Ala	Thr	Asp	Cys	Leu	Val	Asp	Phe	Lys	Lys	Gln	
				380					385					390	
Val	Thr	Val	Gln	Pro	Gly	Ser	Arg	Thr	Arg	Thr	Lys	Ala	Gly	Arg	
				395					400					405	
Gly	Arg	Arg	Arg	Lys	Phe										
				410											

&lt;210&gt; 2

&lt;211&gt; 579

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 945430CD1

&lt;400&gt; 2

Met	Phe	Phe	Arg	Val	Phe	Leu	His	Phe	Ile	Arg	Ser	His	Ser	Ala	
				5					10					15	
Thr	Ala	Val	Asp	Phe	Leu	Pro	Val	Met	Val	His	Arg	Leu	Pro	Val	
				20					25					30	
Phe	Lys	Arg	Tyr	Met	Gly	Asn	Thr	Pro	Gln	Lys	Lys	Ala	Val	Phe	
				35					40					45	
Gly	Gln	Cys	Arg	Gly	Leu	Pro	Cys	Val	Ala	Pro	Leu	Leu	Thr	Thr	
				50					55					60	
Val	Glu	Glu	Ala	Pro	Arg	Gly	Ile	Ser	Ala	Arg	Val	Trp	Gly	His	
				65					70					75	
Phe	Pro	Lys	Trp	Leu	Asn	Gly	Ser	Leu	Leu	Arg	Ile	Gly	Pro	Gly	
				80					85					90	
Lys	Phe	Glu	Phe	Gly	Lys	Asp	Lys	Tyr	Asn	His	Trp	Phe	Asp	Gly	
				95					100					105	
Met	Ala	Leu	Leu	His	Gln	Phe	Arg	Met	Ala	Lys	Gly	Thr	Val	Thr	
				110					115					120	
Tyr	Arg	Ser	Lys	Phe	Leu	Gln	Ser	Asp	Thr	Tyr	Lys	Ala	Asn	Ser	
				125					130					135	
Ala	Lys	Asn	Arg	Ile	Val	Ile	Ser	Glu	Phe	Gly	Thr	Leu	Ala	Leu	
				140					145					150	
Pro	Asp	Pro	Cys	Lys	Asn	Val	Phe	Glu	Arg	Phe	Met	Ser	Arg	Phe	
				155					160					165	

Glu Leu Pro Gly	Lys	Ala	Ala	Ala	Met	Thr	Asp	Asn	Thr	Asn	Val	
	170					175					180	
Asn Tyr Val Arg	Tyr	Lys	Gly	Asp	Tyr	Tyr	Leu	Cys	Thr	Glu	Thr	
	185					190					195	
Asn Phe Met Asn	Lys	Val	Asp	Ile	Glu	Thr	Leu	Glu	Lys	Thr	Glu	
	200					205					210	
Lys Val Asp Trp	Ser	Lys	Phe	Ile	Ala	Val	Asn	Gly	Ala	Thr	Ala	
	215					220					225	
His Pro His Tyr	Asp	Pro	Asp	Gly	Thr	Ala	Tyr	Asn	Met	Gly	Asn	
	230					235					240	
Ser Phe Gly Pro	Tyr	Gly	Phe	Ser	Tyr	Lys	Val	Ile	Arg	Val	Pro	
	245					250					255	
Pro Glu Lys Val	Asp	Leu	Gly	Glu	Thr	Ile	His	Gly	Val	Gln	Val	
	260					265					270	
Ile Cys Ser Ile	Ala	Ser	Thr	Glu	Lys	Gly	Lys	Pro	Ser	Tyr	Tyr	
	275					280					285	
His Ser Phe Gly	Met	Thr	Arg	Asn	Tyr	Ile	Ile	Phe	Ile	Glu	Gln	
	290					295					300	
Pro Leu Lys Met	Asn	Leu	Trp	Lys	Ile	Ala	Thr	Ser	Lys	Ile	Arg	
	305					310					315	
Gly Lys Ala Phe	Ser	Asp	Gly	Ile	Ser	Trp	Glu	Pro	Gln	Cys	Asn	
	320					325					330	
Thr Arg Phe His	Val	Val	Glu	Lys	Arg	Thr	Gly	Gln	Leu	Leu	Pro	
	335					340					345	
Gly Arg Tyr Tyr	Ser	Lys	Pro	Phe	Val	Thr	Phe	His	Gln	Ile	Asn	
	350					355					360	
Ala Phe Glu Asp	Gln	Gly	Cys	Val	Ile	Ile	Asp	Leu	Cys	Ser	Gln	
	365					370					375	
Asp Asn Gly Arg	Thr	Leu	Glu	Val	Tyr	Gln	Leu	Gln	Asn	Leu	Arg	
	380					385					390	
Lys Ala Gly Glu	Gly	Leu	Asp	Gln	Val	His	Asn	Ser	Ala	Ala	Lys	
	395					400					405	
Ser Phe Pro Arg	Arg	Phe	Val	Leu	Pro	Leu	Asn	Val	Ser	Leu	Asn	
	410					415					420	
Ala Pro Glu Gly	Asp	Asn	Leu	Ser	Pro	Leu	Ser	Tyr	Thr	Ser	Ala	
	425					430					435	
Ser Ala Val Lys	Gln	Ala	Asp	Gly	Thr	Ile	Trp	Cys	Ser	His	Glu	
	440					445					450	
Asn Leu His Gln	Glu	Asp	Leu	Glu	Lys	Glu	Gly	Gly	Ile	Glu	Phe	
	455					460					465	
Pro Gln Ile Tyr	Tyr	Asp	Arg	Phe	Ser	Gly	Lys	Lys	Tyr	His	Phe	
	470					475					480	
Phe Tyr Gly Cys	Gly	Phe	Arg	His	Leu	Val	Gly	Asp	Ser	Leu	Ile	
	485					490					495	
Lys Val Asp Val	Val	Asn	Lys	Thr	Leu	Lys	Val	Trp	Arg	Glu	Asp	
	500					505					510	
Gly Phe Tyr Pro	Ser	Glu	Pro	Val	Phe	Val	Pro	Ala	Pro	Gly	Thr	
	515					520					525	
Asn Glu Glu Asp	Gly	Gly	Val	Ile	Leu	Ser	Val	Val	Ile	Thr	Pro	
	530					535					540	
Asn Gln Asn Glu	Ser	Asn	Phe	Leu	Leu	Val	Leu	Asp	Ala	Lys	Asn	
	545					550					555	
Phe Glu Glu Leu	Gly	Arg	Ala	Glu	Val	Pro	Val	Gln	Met	Pro	Tyr	
	560					565					570	
Gly Phe His Gly	Thr	Phe	Ile	Pro	Ile							
	575											

&lt;210&gt; 3

&lt;211&gt; 370

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1305513CD1

&lt;400&gt; 3

```

Met Ala Asn Tyr Ser His Ala Ala Asp Asn Ile Leu Gln Asn Leu
 1          5          10          15
Ser Pro Leu Thr Ala Phe Leu Lys Leu Thr Ser Leu Gly Phe Ile
          20          25          30
Ile Gly Val Ser Val Val Gly Asn Leu Leu Ile Ser Ile Leu Leu
          35          40          45
Val Lys Asp Lys Thr Leu His Arg Ala Pro Tyr Tyr Phe Leu Leu
          50          55          60
Asp Leu Cys Cys Ser Asp Ile Leu Arg Ser Ala Ile Cys Phe Pro
          65          70          75
Phe Val Phe Asn Ser Val Lys Asn Gly Ser Thr Trp Thr Tyr Gly
          80          85          90
Thr Leu Thr Cys Lys Val Ile Ala Phe Leu Gly Val Leu Ser Cys
          95          100          105
Phe His Thr Ala Phe Met Leu Phe Cys Ile Ser Val Thr Arg Tyr
          110          115          120
Leu Ala Ile Ala His His Arg Phe Tyr Thr Lys Arg Leu Thr Phe
          125          130          135
Trp Thr Cys Leu Ala Val Ile Cys Met Val Trp Thr Leu Ser Val
          140          145          150
Ala Met Ala Phe Pro Pro Val Leu Asp Val Gly Thr Tyr Ser Phe
          155          160          165
Ile Arg Glu Lys Asp Gln Cys Thr Phe Gln His Arg Ser Phe Arg
          170          175          180
Ala Asn Asp Ser Leu Gly Phe Met Leu Leu Leu Ala Leu Ile Leu
          185          190          195
Leu Ala Thr Gln Leu Val Tyr Leu Lys Leu Ile Phe Phe Val His
          200          205          210
Asp Arg Arg Lys Met Lys Pro Val Gln Phe Val Ala Ala Val Ser
          215          220          225
Gln Asn Trp Thr Phe His Gly Pro Gly Ala Ser Gly Gln Ala Ala
          230          235          240
Ala Asn Trp Leu Ala Gly Phe Gly Arg Gly Pro Thr Pro Pro Thr
          245          250          255
Leu Leu Gly Ile Arg Gln Asn Ala Asn Thr Thr Gly Arg Arg Arg
          260          265          270
Leu Leu Val Leu Asp Glu Phe Lys Met Glu Lys Arg Ile Ser Arg
          275          280          285
Met Phe Tyr Ile Met Thr Phe Leu Phe Leu Thr Leu Trp Gly Pro
          290          295          300
Tyr Leu Val Ala Cys Tyr Trp Arg Val Phe Ala Arg Gly Pro Val
          305          310          315
Val Pro Gly Gly Phe Leu Thr Ala Ala Val Trp Met Ser Phe Ala
          320          325          330
Gln Ala Gly Ile Asn Pro Phe Val Cys Ile Phe Ser Asn Arg Glu
          335          340          345
Leu Arg Arg Cys Phe Ser Thr Thr Leu Leu Tyr Cys Arg Lys Ser
          350          355          360
Arg Leu Pro Arg Glu Pro Tyr Cys Val Ile
          365          370

```

&lt;210&gt; 4

&lt;211&gt; 267

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1876283CD1

&lt;400&gt; 4

```

Met Ala Pro Trp Ala Leu Leu Ser Pro Gly Val Leu Val Arg Thr
 1          5          10          15
Gly His Thr Val Leu Thr Trp Gly Ile Thr Leu Val Leu Phe Leu

```



	20		25		30
His Asp Thr Glu	Leu Arg Gln Trp Glu	Glu Gln Gly Glu Leu	Leu		
	35		40		45
Leu Pro Leu Thr Phe	Leu Leu Leu Val Leu	Gly Ser Leu Leu Leu			
	50		55		60
Tyr Leu Ala Val Ser	Leu Met Asp Pro Gly	Tyr Val Asn Val Gln			
	65		70		75
Pro Gln Pro Gln Glu	Glu Leu Lys Glu Glu	Gln Thr Ala Met Val			
	80		85		90
Pro Pro Ala Ile Pro	Leu Arg Arg Cys Arg	Tyr Cys Leu Val Leu			
	95		100		105
Gln Pro Leu Arg Ala	Arg His Cys Arg Glu	Cys Arg Arg Cys Val			
	110		115		120
Arg Arg Tyr Asp His	His Cys Pro Trp Met	Glu Asn Cys Val Gly			
	125		130		135
Glu Arg Asn His Pro	Leu Phe Val Val Tyr	Leu Ala Leu Gln Leu			
	140		145		150
Val Val Leu Leu Trp	Gly Leu Tyr Leu Ala	Trp Ser Gly Leu Arg			
	155		160		165
Phe Phe Gln Pro Trp	Gly Leu Trp Leu Arg	Ser Ser Gly Leu Leu			
	170		175		180
Phe Ala Thr Phe Leu	Leu Leu Ser Leu Phe	Ser Leu Val Ala Ser			
	185		190		195
Leu Leu Leu Val Ser	His Leu Tyr Leu Val	Ala Ser Asn Thr Thr			
	200		205		210
Thr Trp Glu Phe Ile	Ser Ser His Arg Ile	Ala Tyr Leu Arg Gln			
	215		220		225
Arg Pro Ser Asn Pro	Phe Asp Arg Gly Leu	Thr Arg Asn Leu Ala			
	230		235		240
His Phe Phe Cys Gly	Trp Pro Ser Gly Ser	Trp Glu Thr Leu Trp			
	245		250		255
Ala Glu Glu Glu Glu	Glu Gly Ser Ser Pro	Ala Val			
	260		265		

&lt;210&gt; 5

&lt;211&gt; 951

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2470285CD1

&lt;400&gt; 5

Met Pro Gly Pro Leu	Gly Leu Leu Cys Phe	Leu Ala Leu Gly Leu		
1	5	10	15	
Leu Gly Ser Ala Gly	Pro Ser Gly Ala Ala	Pro Pro Leu Cys Ala		
	20	25	30	
Ala Pro Cys Ser Cys	Asp Gly Asp Arg Arg	Val Asp Cys Ser Gly		
	35	40	45	
Lys Gly Leu Thr Ala	Val Pro Glu Gly Leu	Ser Ala Phe Thr Gln		
	50	55	60	
Ala Leu Asp Ile Ser	Met Asn Asn Ile Thr	Gln Leu Pro Glu Asp		
	65	70	75	
Ala Phe Lys Asn Phe	Pro Phe Leu Glu Glu	Leu Gln Leu Ala Gly		
	80	85	90	
Asn Asp Leu Ser Phe	Ile His Pro Lys Ala	Leu Ser Gly Leu Lys		
	95	100	105	
Glu Leu Lys Val Leu	Thr Leu Gln Asn Asn	Gln Leu Lys Thr Val		
	110	115	120	
Pro Ser Glu Ala Ile	Arg Gly Leu Ser Ala	Leu Gln Ser Leu Arg		
	125	130	135	
Leu Asp Ala Asn His	Ile Thr Ser Val Pro	Glu Asp Ser Phe Glu		
	140	145	150	
Gly Leu Val Gln Leu	Arg His Leu Trp Leu	Asp Asp Asn Ser Leu		

	155							160					165	
Thr	Glu	Val	Pro	Val	His	Pro	Leu	Ser	Asn	Leu	Pro	Thr	Leu	Gln
				170					175					180
Ala	Leu	Thr	Leu	Ala	Leu	Asn	Lys	Ile	Ser	Ser	Ile	Pro	Asp	Phe
				185					190					195
Ala	Phe	Thr	Asn	Leu	Ser	Ser	Leu	Val	Val	Leu	His	Leu	His	Asn
				200					205					210
Asn	Lys	Ile	Arg	Ser	Leu	Ser	Gln	His	Cys	Phe	Asp	Gly	Leu	Asp
				215					220					225
Asn	Leu	Glu	Thr	Leu	Asp	Leu	Asn	Tyr	Asn	Asn	Leu	Gly	Glu	Phe
				230					235					240
Pro	Gln	Ala	Ile	Lys	Ala	Leu	Pro	Ser	Leu	Lys	Glu	Leu	Gly	Phe
				245					250					255
His	Ser	Asn	Ser	Ile	Ser	Val	Ile	Pro	Asp	Gly	Ala	Phe	Asp	Gly
				260					265					270
Asn	Pro	Leu	Leu	Arg	Thr	Ile	His	Leu	Tyr	Asp	Asn	Pro	Leu	Ser
				275					280					285
Phe	Val	Gly	Asn	Ser	Ala	Phe	His	Asn	Leu	Ser	Asp	Leu	His	Ser
				290					295					300
Leu	Val	Ile	Arg	Gly	Ala	Ser	Met	Val	Gln	Gln	Phe	Pro	Asn	Leu
				305					310					315
Thr	Gly	Thr	Val	His	Leu	Glu	Ser	Leu	Thr	Leu	Thr	Gly	Thr	Lys
				320					325					330
Ile	Ser	Ser	Ile	Pro	Asn	Asn	Leu	Cys	Gln	Glu	Gln	Lys	Met	Leu
				335					340					345
Arg	Thr	Leu	Asp	Leu	Ser	Tyr	Asn	Asn	Ile	Arg	Asp	Leu	Pro	Ser
				350					355					360
Phe	Asn	Gly	Cys	His	Ala	Leu	Glu	Glu	Ile	Ser	Leu	Gln	Arg	Asn
				365					370					375
Gln	Ile	Tyr	Gln	Ile	Lys	Glu	Gly	Thr	Phe	Gln	Gly	Leu	Ile	Ser
				380					385					390
Leu	Arg	Ile	Leu	Asp	Leu	Ser	Arg	Asn	Leu	Ile	His	Glu	Ile	His
				395					400					405
Ser	Arg	Ala	Phe	Ala	Thr	Leu	Gly	Pro	Ile	Thr	Asn	Leu	Asp	Val
				410					415					420
Ser	Phe	Asn	Glu	Leu	Thr	Ser	Phe	Pro	Thr	Glu	Gly	Leu	Asn	Gly
				425					430					435
Leu	Asn	Gln	Leu	Lys	Leu	Val	Gly	Asn	Phe	Lys	Leu	Lys	Glu	Ala
				440					445					450
Leu	Ala	Ala	Lys	Asp	Phe	Val	Asn	Leu	Arg	Ser	Leu	Ser	Val	Pro
				455					460					465
Tyr	Ala	Tyr	Gln	Cys	Cys	Ala	Phe	Trp	Gly	Cys	Asp	Ser	Tyr	Ala
				470					475					480
Asn	Leu	Asn	Thr	Glu	Asp	Asn	Ser	Leu	Gln	Asp	His	Ser	Val	Ala
				485					490					495
Gln	Glu	Lys	Gly	Thr	Ala	Asp	Ala	Ala	Asn	Val	Thr	Ser	Thr	Leu
				500					505					510
Glu	Asn	Glu	Glu	His	Ser	Gln	Ile	Ile	Ile	His	Cys	Thr	Pro	Ser
				515					520					525
Thr	Gly	Ala	Phe	Lys	Pro	Cys	Glu	Tyr	Leu	Leu	Gly	Ser	Trp	Met
				530					535					540
Ile	Arg	Leu	Thr	Val	Trp	Phe	Ile	Phe	Leu	Val	Ala	Leu	Phe	Phe
				545					550					555
Asn	Leu	Leu	Val	Ile	Leu	Thr	Thr	Phe	Ala	Ser	Cys	Thr	Ser	Leu
				560					565					570
Pro	Ser	Ser	Lys	Leu	Phe	Ile	Gly	Leu	Ile	Ser	Val	Ser	Asn	Leu
				575					580					585
Phe	Met	Gly	Ile	Tyr	Thr	Gly	Ile	Leu	Thr	Phe	Leu	Asp	Ala	Val
				590					595					600
Ser	Trp	Gly	Arg	Phe	Ala	Glu	Phe	Gly	Ile	Trp	Trp	Glu	Thr	Gly
				605					610					615
Ser	Gly	Cys	Lys	Val	Ala	Gly	Phe	Leu	Ala	Val	Phe	Ser	Ser	Glu
				620					625					630
Ser	Ala	Ile	Phe	Leu	Leu	Met	Leu	Ala	Thr	Val	Glu	Arg	Ser	Leu

	635		640		645
Ser Ala Lys Asp	Ile Met Lys Asn Gly	Lys Ser Asn His Leu Lys			
	650	655			660
Gln Phe Arg Val	Ala Ala Leu Leu Ala	Phe Leu Gly Ala Thr Val			
	665	670			675
Ala Gly Cys Phe	Pro Leu Phe His Arg	Gly Glu Tyr Ser Ala Ser			
	680	685			690
Pro Leu Cys Leu	Pro Phe Pro Thr Gly	Glu Thr Pro Ser Leu Gly			
	695	700			705
Phe Thr Val Thr	Leu Val Leu Leu Asn	Ser Leu Ala Phe Leu Leu			
	710	715			720
Met Ala Val Ile	Tyr Thr Lys Leu Tyr	Cys Asn Leu Glu Lys Glu			
	725	730			735
Asp Leu Ser Glu	Asn Ser Gln Ser Ser	Met Ile Lys His Val Ala			
	740	745			750
Trp Leu Ile Phe	Thr Asn Cys Ile Phe	Phe Cys Pro Val Ala Phe			
	755	760			765
Phe Ser Phe Ala	Pro Leu Ile Thr Ala	Ile Ser Ile Ser Pro Glu			
	770	775			780
Ile Met Lys Ser	Val Thr Leu Ile Phe	Phe Pro Leu Pro Ala Cys			
	785	790			795
Leu Asn Pro Val	Leu Tyr Val Phe Phe	Asn Pro Lys Phe Lys Glu			
	800	805			810
Asp Trp Lys Leu	Leu Lys Arg Arg Val	Thr Lys Lys Ser Gly Ser			
	815	820			825
Val Ser Val Ser	Ile Ser Ser Gln Gly	Gly Cys Leu Glu Gln Asp			
	830	835			840
Phe Tyr Tyr Asp	Cys Gly Met Tyr Ser	His Leu Gln Gly Asn Leu			
	845	850			855
Thr Val Cys Asp	Cys Cys Glu Ser Phe	Leu Leu Thr Lys Pro Val			
	860	865			870
Ser Cys Lys His	Leu Ile Lys Ser His	Ser Cys Pro Ala Leu Ala			
	875	880			885
Val Ala Ser Cys	Gln Arg Pro Glu Gly	Tyr Trp Ser Asp Cys Gly			
	890	895			900
Thr Gln Ser Ala	His Ser Asp Tyr Ala	Asp Glu Glu Asp Ser Phe			
	905	910			915
Val Ser Asp Ser	Ser Asp Gln Val Gln	Ala Cys Gly Arg Ala Cys			
	920	925			930
Phe Tyr Gln Ser	Arg Gly Phe Pro Leu	Val Arg Tyr Ala Tyr Asn			
	935	940			945
Leu Pro Arg Val	Lys Asp				
	950				

&lt;210&gt; 6

&lt;211&gt; 413

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2925789CD1

&lt;400&gt; 6

Met Gly Thr Phe Cys	Ser Val Ile Lys Phe	Glu Asn Leu Gln Glu			
1	5	10			15
Leu Lys Arg Leu Cys	His Trp Gly Pro	Ile Ile Ala Leu Gly Val			
	20	25			30
Ile Ala Ile Cys Ser	Thr Met Ala Met	Ile Asp Ser Val Leu Trp			
	35	40			45
Tyr Trp Pro Leu His	Thr Thr Gly Gly	Ser Val Asn Phe Ile Met			
	50	55			60
Leu Ile Asn Trp Thr	Val Met Ile Leu Tyr	Asn Tyr Phe Asn Ala			
	65	70			75
Met Phe Val Gly Pro	Gly Phe Val Pro	Leu Gly Trp Lys Pro Glu			

	80		85		90
Ile Ser Gln Asp Thr Met Tyr Leu Gln Tyr Cys Lys Val Cys Gln					
	95		100		105
Ala Tyr Lys Ala Pro Arg Ser His His Cys Arg Lys Cys Asn Arg					
	110		115		120
Cys Val Met Lys Met Asp His His Cys Pro Trp Ile Asn Asn Cys					
	125		130		135
Cys Gly Tyr Gln Asn His Ala Ser Phe Thr Leu Phe Leu Leu Leu					
	140		145		150
Ala Pro Leu Gly Cys Ile His Ala Ala Phe Ile Phe Val Met Thr					
	155		160		165
Met Tyr Thr Gln Leu Tyr His Arg Leu Ser Phe Gly Trp Asn Thr					
	170		175		180
Val Lys Ile Asp Met Ser Ala Ala Arg Arg Asp Pro Leu Pro Ile					
	185		190		195
Val Pro Phe Gly Leu Ala Ala Phe Ala Thr Thr Leu Phe Ala Leu					
	200		205		210
Gly Leu Ala Leu Gly Thr Thr Ile Ala Val Gly Met Leu Phe Phe					
	215		220		225
Ile Gln Met Lys Ile Ile Leu Arg Asn Lys Thr Ser Ile Glu Ser					
	230		235		240
Trp Ile Glu Glu Lys Ala Lys Asp Arg Ile Gln Tyr Tyr Gln Leu					
	245		250		255
Asp Glu Val Phe Val Phe Pro Tyr Asp Met Gly Ser Arg Trp Arg					
	260		265		270
Asn Phe Lys Gln Val Phe Thr Trp Ser Gly Val Pro Glu Gly Asp					
	275		280		285
Gly Leu Glu Trp Pro Val Arg Glu Gly Cys His Gln Tyr Ser Leu					
	290		295		300
Thr Ile Glu Gln Leu Lys Gln Lys Ala Asp Lys Arg Val Arg Ser					
	305		310		315
Val Arg Tyr Lys Val Ile Glu Asp Tyr Ser Gly Ala Cys Cys Pro					
	320		325		330
Leu Asn Lys Gly Ile Lys Thr Phe Phe Thr Ser Pro Cys Thr Glu					
	335		340		345
Glu Pro Arg Ile Gln Leu Gln Lys Gly Glu Phe Ile Leu Ala Thr					
	350		355		360
Arg Gly Leu Arg Tyr Trp Leu Tyr Gly Asp Lys Ile Leu Asp Asp					
	365		370		375
Ser Phe Ile Glu Gly Val Ser Arg Ile Arg Gly Trp Phe Pro Arg					
	380		385		390
Lys Cys Val Glu Lys Cys Pro Cys Asp Ala Glu Thr Asp Gln Ala					
	395		400		405
Pro Glu Gly Glu Lys Lys Asn Arg					
	410				

&lt;210&gt; 7

&lt;211&gt; 144

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3099990CD1

&lt;400&gt; 7

Met Lys Gly Lys Ala Arg Lys Leu Phe Tyr Lys Ala Ile Val Arg					
1	5		10		15
Gly Glu Glu Thr Leu Arg Val Gly Asp Cys Ala Val Phe Leu Ser					
	20		25		30
Ala Gly Arg Pro Asn Leu Pro Tyr Ile Gly Arg Ile Glu Ser Met					
	35		40		45
Trp Glu Ser Trp Gly Ser Asn Met Val Val Lys Val Lys Trp Phe					
	50		55		60
Tyr His Pro Glu Glu Thr Lys Leu Gly Lys Arg Gln Cys Asp Gly					

	65		70		75
Lys Asn Ala Leu Tyr Gln Ser Cys His	Glu Asp Glu Asn Asp Val				
	80		85		90
Gln Thr Ile Ser His Lys Cys Gln Val	Val Ala Arg Glu Gln Tyr				
	95		100		105
Glu Gln Met Ala Arg Ser Arg Lys Cys	Gln Asp Arg Gln Asp Leu				
	110		115		120
Tyr Tyr Leu Ala Gly Thr Tyr Asp Pro	Thr Thr Gly Arg Leu Val				
	125		130		135
Thr Ala Asp Gly Val Pro Ile Leu Cys					
	140				

&lt;210&gt; 8

&lt;211&gt; 174

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 103561CD1

&lt;400&gt; 8

Met Ala Pro Pro Val Arg Leu Glu Arg	Pro Phe Pro Ser Arg Arg				
1	5		10		15
Phe Pro Gly Leu Leu Leu Ala Ala Leu	Val Leu Leu Leu Ser Ser				
	20		25		30
Phe Ser Asp Gln Cys Asn Val Pro Glu	Trp Leu Pro Phe Ala Arg				
	35		40		45
Pro Thr Asn Leu Thr Asp Asp Phe Glu	Phe Pro Ile Gly Thr Tyr				
	50		55		60
Leu Asn Tyr Glu Cys Arg Pro Gly Tyr	Ser Gly Arg Pro Phe Ser				
	65		70		75
Ile Ile Cys Leu Lys Asn Ser Val Trp	Thr Ser Ala Lys Asp Lys				
	80		85		90
Cys Lys Arg Lys Ser Cys Arg Asn Pro	Pro Asp Pro Val Asn Gly				
	95		100		105
Met Ala His Val Ile Lys Asp Ile Gln	Phe Gly Ser Gln Ile Lys				
	110		115		120
Tyr Ser Cys Pro Lys Gly Tyr Arg Leu	Ile Gly Ser Ser Ser Ala				
	125		130		135
Thr Cys Ile Ile Ser Gly Asn Thr Val	Ile Trp Asp Asn Lys Thr				
	140		145		150
Pro Val Cys Asp Ser Glu Leu Lys Tyr	Ala Phe Leu Phe Leu Leu				
	155		160		165
Pro Ile His Ser Asn Phe Ser Leu Glu					
	170				

&lt;210&gt; 9

&lt;211&gt; 449

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 288709CD1

&lt;400&gt; 9

Met Gln Leu Asp Trp Asn Gln Ala Gln	Lys Ser Gly Asp Pro Gly				
1	5		10		15
Pro Ser Val Val Gly Leu Val Ser Ile	Pro Gly Met Gly Lys Leu				
	20		25		30
Leu Ala Glu Ala Pro Leu Val Leu Glu	Pro Glu Lys Gln Met Leu				
	35		40		45
Leu His Glu Thr His Gln Gly Leu Leu	Gln Asp Gly Ser Pro Ile				
	50		55		60
Leu Leu Ser Asp Val Ile Ser Ala Phe	Leu Ser Asn Asn Asp Thr				

	65		70		75
Gln Asn Leu Ser	Ser Pro Val Thr Phe	Thr Phe Ser His Arg	Ser		
	80		85		90
Val Ile Pro Arg	Gln Lys Val Leu Cys	Val Phe Trp Glu His	Gly		
	95		100		105
Gln Asn Gly Cys	Gly His Trp Ala Thr	Thr Gly Cys Ser Thr	Ile		
	110		115		120
Gly Thr Arg Asp	Thr Ser Thr Ile Cys	Arg Cys Thr His Leu	Ser		
	125		130		135
Ser Phe Ala Val	Leu Met Ala His Tyr	Asp Val Gln Glu Glu	Asp		
	140		145		150
Pro Val Leu Thr	Val Ile Thr Tyr Met	Gly Leu Ser Val Ser	Leu		
	155		160		165
Leu Cys Leu Leu	Leu Ala Ala Leu Thr	Phe Leu Leu Cys Lys	Ala		
	170		175		180
Ile Gln Asn Thr	Ser Thr Ser Leu His	Leu Gln Leu Ser Leu	Cys		
	185		190		195
Leu Phe Leu Ala	His Leu Leu Phe Leu	Val Gly Ile Asp Arg	Thr		
	200		205		210
Glu Pro Lys Val	Leu Cys Ser Ile Ile	Ala Gly Ala Leu His	Tyr		
	215		220		225
Leu Tyr Leu Ala	Ala Phe Thr Trp Met	Leu Leu Glu Gly Val	His		
	230		235		240
Leu Phe Leu Thr	Ala Arg Asn Leu Thr	Val Val Asn Tyr Ser	Ser		
	245		250		255
Ile Asn Arg Leu	Met Lys Trp Ile Met	Phe Pro Val Gly Tyr	Gly		
	260		265		270
Val Pro Ala Val	Thr Val Ala Ile Ser	Ala Ala Ser Trp Pro	His		
	275		280		285
Leu Tyr Gly Thr	Ala Asp Arg Cys Trp	Leu His Leu Asp Gln	Gly		
	290		295		300
Phe Met Trp Ser	Phe Leu Gly Pro Val	Cys Ala Ile Phe Ser	Ala		
	305		310		315
Asn Leu Val Leu	Phe Ile Leu Val Phe	Trp Ile Leu Lys Arg	Lys		
	320		325		330
Leu Ser Ser Leu	Asn Ser Glu Val Ser	Thr Ile Gln Asn Thr	Arg		
	335		340		345
Met Leu Ala Phe	Lys Ala Thr Ala Gln	Leu Phe Ile Leu Gly	Cys		
	350		355		360
Thr Trp Cys Leu	Gly Leu Leu Gln Val	Gly Pro Ala Ala Gln	Val		
	365		370		375
Met Ala Tyr Leu	Phe Thr Ile Ile Asn	Ser Leu Gln Gly Phe	Phe		
	380		385		390
Ile Phe Leu Val	Tyr Cys Leu Leu Ser	Gln Gln Val Gln Lys	Gln		
	395		400		405
Tyr Gln Lys Trp	Phe Arg Glu Ile Val	Lys Ser Lys Ser Glu	Ser		
	410		415		420
Glu Thr Tyr Thr	Leu Ser Ser Lys Met	Gly Pro Asp Ser Lys	Pro		
	425		430		435
Ser Glu Gly Asp	Val Phe Pro Gly Gln	Val Lys Arg Lys Tyr			
	440		445		

&lt;210&gt; 10

&lt;211&gt; 126

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 959893CD1

&lt;400&gt; 10

Met Glu Ser Phe Leu Gly Gly Val Leu Leu Ile Leu Trp Leu Gln		
1	5	10 15
Val Asp Trp Val Lys Ser Gln Lys Ile Glu Gln Asn Ser Glu Ala		

	20		25		30
Leu Asn Ile Gln Glu Gly Lys Thr Ala Thr Leu Thr Cys Asn Tyr					
	35		40		45
Thr Asn Tyr Ser Pro Ala Tyr Leu Gln Trp Tyr Arg Gln Asp Pro					
	50		55		60
Gly Arg Gly Pro Val Phe Leu Leu Leu Ile Arg Glu Asn Glu Lys					
	65		70		75
Glu Lys Arg Lys Glu Arg Leu Lys Val Thr Phe Asp Thr Thr Leu					
	80		85		90
Lys Gln Ser Leu Phe His Ile Thr Ala Ser Gln Pro Ala Asp Ser					
	95		100		105
Ala Asn Tyr Leu Cys Ala Leu Gly Gly Arg Gly Thr Asn Ser Pro					
	110		115		120
Leu Gly Gln Ala Leu Ser					
	125				

&lt;210&gt; 11

&lt;211&gt; 273

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1414179CD1

&lt;400&gt; 11

Met Gly Arg Ser Arg Ser Arg Ser Ser Ser Arg Ser Lys His Thr					
1	5		10		15
Lys Ser Ser Lys His Asn Lys Lys Arg Ser Arg Ser Arg Ser Arg					
	20		25		30
Ser Arg Asp Lys Glu Arg Val Arg Lys Arg Ser Lys Ser Arg Glu					
	35		40		45
Ser Lys Arg Asn Arg Arg Arg Glu Ser Arg Ser Arg Ser Arg Ser					
	50		55		60
Thr Asn Thr Ala Val Ser Arg Arg Glu Arg Asp Arg Glu Arg Ala					
	65		70		75
Ser Ser Pro Pro Asp Arg Ile Asp Ile Phe Gly Arg Thr Val Ser					
	80		85		90
Lys Arg Ser Ser Leu Asp Glu Lys Gln Lys Arg Glu Glu Glu Glu					
	95		100		105
Lys Lys Ala Glu Phe Glu Arg Gln Arg Lys Ile Arg Gln Gln Glu					
	110		115		120
Ile Glu Glu Lys Leu Ile Glu Glu Glu Thr Ala Arg Arg Val Glu					
	125		130		135
Glu Leu Val Ala Lys Arg Val Glu Glu Glu Leu Glu Lys Arg Lys					
	140		145		150
Asp Glu Ile Glu Arg Glu Val Leu Arg Arg Val Glu Glu Ala Lys					
	155		160		165
Arg Ile Met Glu Lys Gln Leu Leu Glu Glu Leu Glu Arg Gln Arg					
	170		175		180
Gln Ala Glu Leu Ala Ala Gln Lys Ala Arg Glu Glu Glu Glu Arg					
	185		190		195
Ala Lys Arg Glu Glu Leu Glu Arg Ile Leu Glu Glu Asn Asn Arg					
	200		205		210
Lys Ile Ala Glu Ala Gln Ala Lys Leu Ala Glu Glu Gln Leu Arg					
	215		220		225
Ile Val Glu Glu Gln Arg Lys Ile His Glu Glu Arg Met Lys Leu					
	230		235		240
Glu Gln Glu Arg Gln Arg Gln Gln Lys Glu Glu Gln Lys Ile Ile					
	245		250		255
Leu Gly Lys Gly Lys Ser Arg Pro Lys Leu Ser Phe Ser Leu Lys					
	260		265		270
Thr Gln Asp					

&lt;210&gt; 12

<211> 140  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2197211CD1

<400> 12  
 Met Glu Lys Met Leu Glu Cys Ala Phe Ile Val Leu Trp Leu Gln  
 1 5 10 15  
 Leu Gly Trp Leu Ser Gly Glu Asp Gln Val Thr Gln Ser Pro Glu  
 20 25 30  
 Ala Leu Arg Leu Gln Glu Gly Glu Ser Ser Leu Asn Cys Ser  
 35 40 45  
 Tyr Thr Val Ser Gly Leu Arg Gly Leu Phe Trp Tyr Arg Gln Asp  
 50 55 60  
 Pro Gly Lys Gly Pro Glu Phe Leu Phe Thr Leu Tyr Ser Ala Gly  
 65 70 75  
 Glu Glu Lys Glu Lys Glu Arg Leu Lys Ala Thr Leu Thr Lys Lys  
 80 85 90  
 Glu Ser Phe Leu His Ile Thr Ala Pro Lys Pro Glu Asp Ser Ala  
 95 100 105  
 Ser Tyr Leu Cys Ala Val Gln Gly Gly Ile Gly Asn Val Leu His  
 110 115 120  
 Cys Gly Ser Gly Thr Gln Val Val Val Leu Pro His Ile Arg Asp  
 125 130 135  
 Pro Gly Pro Ala Val  
 140

<210> 13  
 <211> 479  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2263653CD1

<400> 13  
 Met Ala Val Leu Gly Val Gln Leu Val Val Thr Leu Leu Thr Ala  
 1 5 10 15  
 Thr Leu Met His Arg Leu Ala Pro His Cys Ser Phe Ala Arg Trp  
 20 25 30  
 Leu Leu Cys Asn Gly Ser Leu Phe Arg Tyr Lys His Pro Ser Glu  
 35 40 45  
 Glu Glu Leu Arg Ala Leu Ala Gly Lys Pro Arg Pro Arg Gly Arg  
 50 55 60  
 Lys Glu Arg Trp Ala Asn Gly Leu Ser Glu Glu Lys Pro Leu Ser  
 65 70 75  
 Val Pro Arg Asp Ala Pro Phe Gln Leu Glu Thr Cys Pro Leu Thr  
 80 85 90  
 Thr Val Asp Ala Leu Val Leu Arg Phe Phe Leu Glu Tyr Gln Trp  
 95 100 105  
 Phe Val Asp Phe Ala Val Tyr Ser Gly Gly Val Tyr Leu Phe Thr  
 110 115 120  
 Glu Ala Tyr Tyr Tyr Met Leu Gly Pro Ala Lys Glu Thr Asn Ile  
 125 130 135  
 Ala Val Phe Trp Cys Leu Leu Thr Val Thr Phe Ser Ile Lys Met  
 140 145 150  
 Phe Leu Thr Val Thr Arg Leu Tyr Phe Ser Ala Glu Glu Gly Gly  
 155 160 165  
 Glu Arg Ser Val Cys Leu Thr Phe Ala Phe Leu Phe Leu Leu Leu  
 170 175 180  
 Ala Met Leu Val Gln Val Val Arg Glu Glu Thr Leu Glu Leu Gly



Leu Glu Pro Gly	185	Leu Ala Ser Met Thr	190	Gln Asn Leu Glu Pro	195
	200		205		210
Leu Lys Lys Gln	215	Gly Trp Asp Trp Ala	220	Leu Pro Val Ala Lys	225
	230		235		240
Ala Ile Arg Val	245	Gly Leu Ala Val Val	250	Gly Ser Val Leu Gly	255
	260		265		270
Phe Leu Thr Phe	275	Pro Gly Leu Arg Leu	280	Ala Gln Thr His Arg	285
	290		295		300
Ala Leu Thr Met	305	Ser Glu Asp Arg Pro	310	Met Leu Gln Phe Leu	315
	320		325		330
His Thr Ser Phe	335	Leu Ser Pro Leu Phe	340	Ile Leu Trp Leu Trp	345
	350		355		360
Lys Pro Ile Ala	365	Arg Asp Phe Leu His	370	Gln Pro Pro Phe Gly	375
	380		385		390
Thr Arg Phe Ser	395	Leu Leu Ser Asp Ser	400	Ala Phe Asp Ser Gly	405
	410		415		420
Leu Trp Leu Leu	425	Val Val Leu Cys Leu	430	Leu Arg Leu Ala Val	435
	440		445		450
Arg Pro His Leu	455	Gln Ala Tyr Leu Cys	460	Leu Ala Lys Ala Arg	465
	470		475		
Glu Gln Leu Arg		Arg Glu Ala Gly Arg		Ile Glu Ala Arg Glu	
Gln Gln Arg Val		Val Arg Val Tyr Cys		Tyr Val Thr Val Val	
Leu Gln Tyr Leu		Thr Pro Leu Ile Leu		Thr Leu Asn Cys Thr	
Leu Leu Lys Thr		Leu Gly Gly Tyr Ser		Trp Gly Leu Gly Pro	
Pro Leu Leu Ser		Pro Asp Pro Ser Ser		Ala Ser Ala Ala Pro	
Gly Ser Gly Glu		Asp Glu Val Gln Gln		Thr Ala Ala Arg Ile	
Gly Ala Leu Gly		Gly Leu Leu Thr Pro		Leu Phe Leu Arg Gly	
Leu Ala Tyr Leu		Ile Trp Trp Thr Ala		Ala Cys Gln Leu Leu	
Ser Leu Phe Gly		Leu Tyr Phe His Gln		His Leu Ala Gly Ser	

&lt;210&gt; 14

&lt;211&gt; 99

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2504590CD1

&lt;400&gt; 14

Met Pro Arg Leu Lys	5	Asp Pro Phe Phe Cys	10	Tyr Gln Met Glu Ser	15
His Cys Val Pro Arg	20	Leu Glu Cys Ser Gly	25	Ala Ile Ser Thr His	30
Cys Lys Leu Cys Leu	35	Pro Gly Ser Arg His	40	Ser Pro Ala Ser Gly	45
Ser Arg Val Ala Gly	50	Thr Thr Gly Ala Arg	55	His His Ser Trp Leu	60
Ile Leu Phe Val Phe	65	Ser Val Glu Thr Gly	70	Tyr His His Val Ser	75
Gln Asp Gly Leu Asp	80	Leu Pro Asp Leu Val	85	Ile Arg Pro Pro Gln	90
Ser Pro Lys Val Leu	95	Gly Leu Gln Ala			

&lt;210&gt; 15

<211> 349  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2529619CD1

<400> 15

Met	Ser	Ser	Glu	Met	Val	Lys	Asn	Gln	Thr	Met	Val	Thr	Glu	Phe
1				5					10					15
Leu	Leu	Leu	Gly	Phe	Leu	Leu	Gly	Pro	Arg	Ile	Gln	Met	Leu	Leu
			20						25					30
Phe	Gly	Leu	Phe	Ser	Leu	Phe	Tyr	Val	Phe	Thr	Leu	Leu	Gly	Asn
			35						40					45
Gly	Thr	Ile	Leu	Gly	Leu	Ile	Ser	Leu	Asp	Ser	Arg	Leu	His	Thr
			50						55					60
Pro	Met	Tyr	Phe	Phe	Leu	Ser	His	Leu	Ala	Val	Val	Asn	Ile	Ala
			65						70					75
Tyr	Ala	Cys	Asn	Thr	Val	Pro	Gln	Met	Leu	Val	Asn	Leu	Leu	His
			80						85					90
Pro	Ala	Lys	Pro	Ile	Ser	Phe	Ala	Gly	Cys	Met	Thr	Thr	Thr	Phe
			95						100					105
Leu	Phe	Leu	Ser	Phe	Ala	His	Thr	Glu	Cys	Leu	Leu	Leu	Val	Leu
			110						115					120
Met	Ser	Tyr	Asp	Arg	Tyr	Val	Ala	Ile	Cys	His	Pro	Leu	Arg	Tyr
			125						130					135
Phè	Ile	Ile	Met	Thr	Trp	Lys	Val	Cys	Ile	Thr	Leu	Ala	Ile	Thr
			140						145					150
Ser	Trp	Thr	Cys	Gly	Ser	Leu	Leu	Ala	Met	Val	His	Val	Ser	Leu
			155						160					165
Ile	Leu	Arg	Leu	Pro	Phe	Cys	Gly	Pro	Arg	Glu	Ile	Asn	His	Phe
			170						175					180
Phe	Cys	Glu	Ile	Leu	Ser	Val	Leu	Arg	Leu	Ala	Cys	Ala	Asp	Thr
			185						190					195
Trp	Leu	Asn	Gln	Val	Val	Ile	Phe	Ala	Ala	Cys	Met	Phe	Ile	Leu
			200						205					210
Val	Gly	Pro	Leu	Cys	Leu	Val	Leu	Val	Ser	Tyr	Ser	His	Ile	Leu
			215						220					225
Ala	Ala	Ile	Leu	Arg	Ile	Gln	Ser	Gly	Glu	Gly	Arg	Arg	Lys	Ala
			230						235					240
Phe	Ser	Thr	Cys	Ser	Ser	His	Leu	Cys	Val	Val	Gly	Leu	Phe	Phe
			245						250					255
Gly	Ser	Ala	Ile	Val	Met	Tyr	Met	Ala	Pro	Lys	Ser	Arg	His	Pro
			260						265					270
Glu	Glu	Gln	Gln	Lys	Val	Leu	Phe	Leu	Phe	Tyr	Ser	Ser	Phe	Asn
			275						280					285
Pro	Met	Leu	Asn	Pro	Leu	Ile	Tyr	Asn	Leu	Arg	Asn	Val	Glu	Val
			290						295					300
Arg	Cys	Pro	Glu	Glu	Ser	Thr	Val	Gln	Glu	Lys	Ser	Phe	Leu	Arg
			305						310					315
Gly	Val	Thr	Phe	Glu	Leu	Pro	Ala	Ser	Val	Val	Thr	Trp	Thr	Leu
			320						325					330
Asp	Ala	Gln	Leu	Leu	Pro	Gln	Ser	Arg	Lys	Val	Tyr	Phe	Ser	Leu
			335						340					345

Ser Val Leu Tyr

<210> 16  
 <211> 373  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 5467661CD1

&lt;400&gt; 16

```

Met Asp Thr Leu Glu Glu Val Thr Trp Ala Asn Gly Ser Thr Ala
 1          5          10          15
Leu Pro Pro Pro Leu Ala Pro Asn Ile Ser Val Pro His Arg Cys
 20          25          30
Leu Leu Leu Leu Tyr Glu Asp Ile Gly Thr Ser Arg Val Arg Tyr
 35          40          45
Trp Asp Leu Leu Leu Ile Pro Asn Val Leu Phe Leu Ile Phe
 50          55          60
Leu Leu Trp Lys Leu Pro Ser Ala Arg Ala Lys Ile Arg Ile Thr
 65          70          75
Ser Ser Pro Ile Phe Ile Thr Phe Tyr Ile Leu Val Phe Val Val
 80          85          90
Ala Leu Val Gly Ile Ala Arg Ala Val Val Ser Met Thr Val Ser
 95          100          105
Thr Ser Asn Ala Ala Thr Val Ala Asp Lys Ile Leu Trp Glu Ile
 110          115          120
Thr Arg Phe Phe Leu Leu Ala Ile Glu Leu Ser Val Ile Ile Leu
 125          130          135
Gly Leu Ala Phe Gly His Leu Glu Ser Lys Ser Ser Ile Lys Arg
 140          145          150
Val Leu Ala Ile Thr Thr Val Leu Ser Leu Ala Tyr Ser Val Thr
 155          160          165
Gln Gly Thr Leu Glu Ile Leu Tyr Pro Asp Ala His Leu Ser Ala
 170          175          180
Glu Asp Phe Asn Ile Tyr Gly His Gly Gly Arg Gln Phe Trp Leu
 185          190          195
Val Ser Ser Cys Phe Phe Phe Leu Val Tyr Ser Leu Val Val Ile
 200          205          210
Leu Pro Lys Thr Pro Leu Lys Glu Arg Ile Ser Leu Pro Ser Arg
 215          220          225
Arg Ser Phe Tyr Val Tyr Ala Gly Ile Leu Ala Leu Leu Asn Leu
 230          235          240
Leu Gln Gly Leu Gly Ser Val Leu Leu Cys Phe Asp Ile Ile Glu
 245          250          255
Gly Leu Cys Cys Val Asp Ala Thr Thr Phe Leu Tyr Phe Ser Phe
 260          265          270
Phe Ala Pro Leu Ile Tyr Val Ala Phe Leu Arg Gly Phe Phe Gly
 275          280          285
Ser Glu Pro Lys Ile Leu Phe Ser Tyr Lys Cys Gln Val Asp Glu
 290          295          300
Thr Glu Glu Pro Asp Val His Leu Pro Gln Pro Tyr Ala Val Ala
 305          310          315
Arg Arg Glu Gly Leu Glu Ala Ala Gly Ala Ala Gly Ala Ser Ala
 320          325          330
Ala Ser Tyr Ser Ser Thr Gln Phe Asp Ser Ala Gly Gly Val Ala
 335          340          345
Tyr Leu Asp Asp Ile Ala Ser Met Pro Cys His Thr Gly Ser Ile
 350          355          360
Asn Ser Thr Asp Ser Glu Arg Trp Lys Ala Ile Asn Ala
 365          370

```

&lt;210&gt; 17

&lt;211&gt; 353

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 229740CD1

&lt;400&gt; 17

```

Met Leu Lys Met Met Glu Val Tyr Lys Glu Pro Arg Glu Gln Pro

```

1	5	10	15
Ile Phe Thr Thr Arg	Ala His Val Phe	Gln Ile Asp Pro Asn	Thr
20	25	30	
Lys Lys Asn Trp Met	Pro Ala Ser Lys	Gln Ala Val Thr Val	Ser
35	40	45	
Tyr Phe Tyr Asp Val	Thr Arg Asn Ser	Tyr Arg Ile Ile Ser	Val
50	55	60	
Asp Gly Ala Lys Val	Ile Ile Asn Ser	Thr Ile Thr Pro Asn	Met
65	70	75	
Thr Phe Thr Asn Thr	Ser Gln Thr Ser	Gly Gln Trp Ala Asp	Ser
80	85	90	
Arg Ala Asn Thr Val	Phe Gly Leu Gly	Phe Ser Ser Glu Gln	Gln
95	100	105	
Leu Thr Lys Phe Ala	Glu Lys Phe Gln	Glu Val Lys Glu Ala	Ala
110	115	120	
Lys Ile Ala Lys Asp	Lys Thr Gln Glu	Lys Ile Glu Thr Ser	Ser
125	130	135	
Asn His Ser Gln Ala	Ser Ser Val Asn	Gly Thr Asp Asp Glu	Lys
140	145	150	
Ala Ser His Ala Gly	Pro Ala Asn Thr	His Leu Lys Ser Glu	Asn
155	160	165	
Asp Lys Leu Lys Ile	Ala Leu Thr Gln	Ser Ala Ala Asn Val	Lys
170	175	180	
Lys Trp Glu Ile Glu	Leu Gln Thr Leu	Arg Glu Ser Asn Ala	Arg
185	190	195	
Leu Thr Thr Ala Leu	Gln Glu Ser Ala	Ala Ser Val Glu Gln	Trp
200	205	210	
Lys Arg Gln Phe Ser	Ile Cys His Asp	Glu Asn Asp Gln Leu	Arg
215	220	225	
Asn Lys Ile Asp Glu	Leu Glu Glu Gln	Cys Ser Glu Ile Asn	Arg
230	235	240	
Glu Lys Glu Lys Asn	Thr Gln Leu Lys	Arg Arg Ile Glu Glu	Leu
245	250	255	
Glu Ala Glu Leu Arg	Glu Lys Glu Thr	Glu Leu Lys Asp Leu	Arg
260	265	270	
Lys Gln Ser Glu Ile	Ile Pro Gln Leu	Met Ser Glu Cys Glu	Tyr
275	280	285	
Val Ser Glu Lys Leu	Glu Ala Ala Glu	Arg Asp Asn Gln Asn	Leu
290	295	300	
Glu Asp Lys Val Arg	Ser Leu Lys Thr	Asp Ile Glu Glu Ser	Lys
305	310	315	
Tyr Arg Gln Arg His	Leu Lys Val Glu	Leu Lys Ser Phe Leu	Glu
320	325	330	
Val Leu Asp Gly Lys	Ile Asp Asp Leu	His Asp Phe Arg Arg	Gly
335	340	345	
Leu Ser Lys Leu Gly	Thr Asp Asn		
350			

&lt;210&gt; 18

&lt;211&gt; 441

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1317467CD1

&lt;400&gt; 18

Met Leu Leu Pro Gly	Arg Ala Arg Gln	Pro Pro Thr Pro	Gln Pro
1	5	10	15
Val Gln His Pro Gly	Leu Arg Arg Gln	Val Glu Pro Pro	Gly Gln
20	25	30	
Leu Leu Arg Leu Phe	Tyr Cys Thr Val	Leu Val Cys Ser	Lys Glu
35	40	45	
Ile Ser Ala Leu Thr	Asp Phe Ser Gly	Tyr Leu Thr Lys	Leu Leu

Gln	Asn	His	Thr	Thr	Tyr	Ala	Cys	Asp	Gly	Asp	Tyr	Leu	Asn	Leu	50	55	60
				65					70								75
Gln	Cys	Pro	Arg	His	Ser	Thr	Ile	Ser	Val	Gln	Ser	Ala	Phe	Tyr			
				80					85								90
Gly	Gln	Asp	Tyr	Gln	Met	Cys	Ser	Ser	Gln	Lys	Pro	Ala	Ser	Gln			
				95					100								105
Arg	Glu	Asp	Ser	Leu	Thr	Cys	Val	Ala	Ala	Thr	Thr	Phe	Gln	Lys			
				110					115								120
Val	Leu	Asp	Glu	Cys	Gln	Asn	Gln	Arg	Ala	Cys	His	Leu	Leu	Val			
				125					130								135
Asn	Ser	Arg	Val	Phe	Gly	Pro	Asp	Leu	Cys	Pro	Gly	Ser	Ser	Lys			
				140					145								150
Tyr	Leu	Leu	Val	Ser	Phe	Lys	Cys	Gln	Pro	Asn	Glu	Leu	Lys	Asn			
				155					160								165
Lys	Thr	Val	Cys	Glu	Asp	Gln	Glu	Leu	Lys	Leu	His	Cys	His	Glu			
				170					175								180
Ser	Lys	Phe	Leu	Asn	Ile	Tyr	Ser	Ala	Thr	Tyr	Gly	Arg	Arg	Thr			
				185					190								195
Gln	Glu	Arg	Asp	Ile	Cys	Ser	Ser	Lys	Ala	Glu	Arg	Leu	Pro	Pro			
				200					205								210
Phe	Asp	Cys	Leu	Ser	Tyr	Ser	Ala	Leu	Gln	Val	Leu	Ser	Arg	Arg			
				215					220								225
Cys	Tyr	Gly	Lys	Gln	Arg	Cys	Lys	Ile	Ile	Val	Asn	Asn	His	His			
				230					235								240
Phe	Gly	Ser	Pro	Cys	Leu	Pro	Gly	Val	Lys	Lys	Tyr	Leu	Thr	Val			
				245					250								255
Thr	Tyr	Ala	Cys	Val	Pro	Lys	Asn	Ile	Leu	Thr	Ala	Ile	Asp	Pro			
				260					265								270
Ala	Ile	Ala	Asn	Leu	Lys	Pro	Ser	Leu	Lys	Gln	Lys	Asp	Gly	Glu			
				275					280								285
Tyr	Gly	Ile	Asn	Phe	Asp	Pro	Ser	Gly	Ser	Lys	Val	Leu	Arg	Lys			
				290					295								300
Asp	Gly	Ile	Leu	Val	Ser	Asn	Ser	Leu	Ala	Ala	Phe	Ala	Tyr	Ile			
				305					310								315
Arg	Ala	His	Pro	Glu	Arg	Ala	Ala	Leu	Leu	Phe	Val	Ser	Ser	Val			
				320					325								330
Cys	Ile	Gly	Leu	Ala	Leu	Thr	Leu	Cys	Ala	Leu	Val	Ile	Arg	Glu			
				335					340								345
Ser	Cys	Ala	Lys	Asp	Phe	Arg	Asp	Leu	Gln	Leu	Gly	Arg	Glu	Gln			
				350					355								360
Leu	Val	Pro	Gly	Ser	Asp	Lys	Val	Glu	Glu	Asp	Ser	Glu	Asp	Glu			
				365					370								375
Glu	Glu	Glu	Glu	Asp	Pro	Ser	Glu	Ser	Asp	Phe	Pro	Gly	Glu	Leu			
				380					385								390
Ser	Gly	Phe	Cys	Arg	Thr	Ser	Tyr	Pro	Ile	Tyr	Ser	Ser	Ile	Glu			
				395					400								405
Ala	Ala	Glu	Leu	Ala	Glu	Arg	Ile	Glu	Arg	Arg	Glu	Gln	Ile	Ile			
				410					415								420
Gln	Glu	Ile	Trp	Met	Asn	Ser	Gly	Leu	Asp	Thr	Ser	Leu	Pro	Arg			
				425					430								435
Asn	Met	Gly	Gln	Phe	Tyr												
				440													

&lt;210&gt; 19

&lt;211&gt; 310

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2279267CD1

&lt;400&gt; 19

Met Gly Asp Asn Ile Thr Ser Ile Thr Glu Phe Leu Leu Leu Gly

1	5	10	15
Phe Pro Val Gly	Pro Arg Ile Gln Met	Leu Leu Phe Gly Leu	Phe
	20	25	30
Ser Leu Phe Tyr	Val Phe Thr Leu Leu	Gly Asn Gly Thr Ile	Leu
	35	40	45
Gly Leu Ile Ser	Leu Asp Ser Arg Leu	His Ala Pro Met Tyr	Phe
	50	55	60
Phe Leu Ser His	Leu Ala Val Val Asp	Ile Ala Tyr Ala Cys	Asn
	65	70	75
Thr Val Pro Arg	Met Leu Val Asn Leu	Leu His Pro Ala Lys	Pro
	80	85	90
Ile Ser Phe Ala	Gly Arg Met Met Gln	Thr Phe Leu Phe Ser	Thr
	95	100	105
Phe Ala Val Thr	Glu Cys Leu Leu Leu	Val Val Met Ser Tyr	Asp
	110	115	120
Leu Tyr Val Ala	Ile Cys His Pro Leu	Arg Tyr Leu Ala Ile	Met
	125	130	135
Thr Trp Arg Val	Cys Ile Thr Leu Ala	Val Thr Ser Trp Thr	Thr
	140	145	150
Gly Val Leu Leu	Ser Leu Ile His Leu	Val Leu Leu Leu Pro	Leu
	155	160	165
Pro Phe Cys Arg	Pro Gln Lys Ile Tyr	His Phe Phe Cys Glu	Ile
	170	175	180
Leu Ala Val Leu	Lys Leu Ala Cys Ala	Asp Thr His Ile Asn	Glu
	185	190	195
Asn Met Val Leu	Ala Gly Ala Ile Ser	Gly Leu Val Gly Pro	Leu
	200	205	210
Ser Thr Ile Val	Val Ser Tyr Met Cys	Ile Leu Cys Ala Ile	Leu
	215	220	225
Gln Ile Gln Ser	Arg Glu Val Gln Arg	Lys Ala Phe Cys Thr	Cys
	230	235	240
Phe Ser His Leu	Cys Val Ile Gly Leu	Phe Tyr Gly Thr Ala	Ile
	245	250	255
Ile Met Tyr Val	Gly Pro Arg Tyr Gly	Asn Pro Lys Glu Gln	Lys
	260	265	270
Lys Tyr Leu Leu	Leu Phe His Ser Leu	Phe Asn Pro Met Leu	Asn
	275	280	285
Pro Leu Ile Cys	Ser Leu Arg Asn Ser	Glu Val Lys Asn Thr	Leu
	290	295	300
Lys Arg Val Leu	Gly Val Glu Arg Ala	Leu	
	305	310	

&lt;210&gt; 20

&lt;211&gt; 438

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2436258CD1

&lt;400&gt; 20

Met Glu Val Gly	Gly Asp Thr Ala Ala	Pro Ala Pro Gly Gly	Ala
1	5	10	15
Glu Asp Leu Glu	Asp Thr Gln Phe Pro	Ser Glu Glu Ala Arg	Glu
	20	25	30
Gly Gly Gly Val	His Ala Val Pro Pro	Asp Pro Glu Asp Glu	Gly
	35	40	45
Leu Glu Glu Thr	Glu Asp His Lys Leu	Val Phe Leu Gln Gln	Gly
	50	55	60
Pro Leu Leu Leu	Val Ala Met Ser Arg	Thr Ser Gln Ser Ala	Ala
	65	70	75
Gln Leu Arg Gly	Glu Leu Leu Ala Val	His Ala Gln Ile Val	Ser
	80	85	90
Thr Leu Thr Arg	Ala Ser Val Ala Arg	Ile Phe Ala His Lys	Gln

Asn	Tyr	Asp	Leu	Arg	Arg	Leu	Leu	Ala	Gly	Ser	Glu	Arg	Thr	Leu	105
															100
Asp	Arg	Leu	Leu	Asp	Ser	Met	Glu	Gln	Asp	Pro	Gly	Ala	Leu	Leu	115
															110
Leu	Gly	Ala	Val	Arg	Cys	Val	Pro	Leu	Ala	Arg	Pro	Leu	Arg	Asp	125
															120
Ala	Leu	Gly	Ala	Leu	Leu	Arg	Arg	Cys	Thr	Ala	Pro	Gly	Leu	Ala	130
															135
Leu	Ser	Val	Leu	Ala	Val	Gly	Gly	Arg	Leu	Ile	Thr	Ala	Ala	Gln	140
															145
Glu	Arg	Asn	Val	Leu	Ala	Glu	Cys	Arg	Leu	Asp	Pro	Ala	Asp	Leu	155
															160
Gln	Leu	Leu	Leu	Asp	Trp	Val	Gly	Ala	Pro	Ala	Phe	Ala	Ala	Gly	170
															175
Glu	Ala	Trp	Ala	Pro	Val	Cys	Leu	Pro	Arg	Phe	Asn	Pro	Asp	Gly	185
															190
Phe	Phe	Tyr	Ala	Tyr	Val	Ala	Arg	Leu	Asp	Ala	Met	Pro	Val	Cys	200
															205
Leu	Leu	Leu	Leu	Gly	Thr	Gln	Arg	Glu	Ala	Phe	His	Ala	Met	Ala	215
															220
Ala	Cys	Arg	Arg	Leu	Val	Glu	Asp	Gly	Met	His	Ala	Leu	Gly	Ala	230
															235
Met	Arg	Ala	Leu	Gly	Glu	Ala	Ala	Ser	Phe	Ser	Asn	Ala	Ser	Ser	245
															250
Ala	Ser	Ala	Pro	Ala	Tyr	Ser	Val	Gln	Ala	Val	Gly	Ala	Pro	Gly	260
															265
Leu	Arg	His	Phe	Leu	Tyr	Lys	Pro	Leu	Asp	Ile	Pro	Asp	His	His	275
															280
Arg	Gln	Leu	Pro	Gln	Phe	Thr	Ser	Pro	Glu	Leu	Glu	Ala	Pro	Tyr	290
															295
Ser	Arg	Glu	Glu	Glu	Arg	Gln	Arg	Leu	Ser	Asp	Leu	Tyr	His	Arg	305
															310
Leu	His	Ala	Arg	Leu	His	Ser	Thr	Ser	Arg	Pro	Leu	Arg	Leu	Ile	320
															325
Tyr	His	Val	Ala	Glu	Lys	Glu	Thr	Leu	Leu	Ala	Trp	Val	Thr	Ser	335
															340
Lys	Phe	Glu	Leu	Tyr	Thr	Cys	Leu	Ser	Pro	Leu	Val	Thr	Lys	Ala	350
															355
Gly	Ala	Ile	Leu	Val	Val	Thr	Lys	Leu	Leu	Arg	Trp	Val	Lys	Lys	365
															370
Glu	Glu	Asp	Arg	Leu	Phe	Ile	Arg	Tyr	Pro	Pro	Lys	Tyr	Ser	Thr	380
															385
Pro	Pro	Ala	Thr	Ser	Thr	Asp	Gln	Ala	Ala	His	Asn	Gly	Leu	Phe	395
															400

Thr Gly Leu

<210> 21

<211> 357

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2681738CD1

<400> 21

Met	Ala	Thr	Thr	Val	Pro	Asp	Gly	Cys	Arg	Asn	Gly	Leu	Lys	Ser	
1				5					10					15	
Lys	Tyr	Tyr	Arg	Leu	Cys	Asp	Lys	Ala	Glu	Ala	Trp	Gly	Ile	Val	
				20					25					30	
Leu	Glu	Thr	Val	Ala	Thr	Ala	Gly	Val	Val	Thr	Ser	Val	Ala	Phe	
				35					40					45	
Met	Leu	Thr	Leu	Pro	Ile	Leu	Val	Cys	Lys	Val	Gln	Asp	Ser	Asn	

				50					55					60
Arg	Arg	Lys	Met	Leu	Pro	Thr	Gln	Phe	Leu	Phe	Leu	Leu	Gly	Val
				65					70					75
Leu	Gly	Ile	Phe	Gly	Leu	Thr	Phe	Ala	Phe	Ile	Ile	Gly	Leu	Asp
				80					85					90
Gly	Ser	Thr	Gly	Pro	Thr	Arg	Phe	Phe	Leu	Phe	Gly	Ile	Leu	Phe
				95					100					105
Ser	Ile	Cys	Phe	Ser	Cys	Leu	Leu	Ala	His	Ala	Val	Ser	Leu	Thr
				110					115					120
Lys	Leu	Val	Arg	Gly	Arg	Lys	Pro	Leu	Ser	Leu	Leu	Val	Ile	Leu
				125					130					135
Gly	Leu	Ala	Val	Gly	Phe	Ser	Leu	Val	Gln	Asp	Val	Ile	Ala	Ile
				140					145					150
Glu	Tyr	Ile	Val	Leu	Thr	Met	Asn	Arg	Thr	Asn	Val	Asn	Val	Phe
				155					160					165
Ser	Glu	Leu	Ser	Ala	Pro	Arg	Arg	Asn	Glu	Asp	Phe	Val	Leu	Leu
				170					175					180
Leu	Thr	Tyr	Val	Leu	Phe	Leu	Met	Ala	Leu	Thr	Phe	Leu	Met	Ser
				185					190					195
Ser	Phe	Thr	Phe	Cys	Gly	Ser	Phe	Thr	Gly	Trp	Lys	Arg	His	Gly
				200					205					210
Ala	His	Ile	Tyr	Leu	Thr	Met	Leu	Leu	Ser	Ile	Ala	Ile	Trp	Val
				215					220					225
Ala	Trp	Ile	Thr	Leu	Leu	Met	Leu	Pro	Asp	Phe	Asp	Arg	Arg	Trp
				230					235					240
Asp	Asp	Thr	Ile	Leu	Ser	Ser	Ala	Leu	Ala	Ala	Asn	Gly	Trp	Val
				245					250					255
Phe	Leu	Leu	Ala	Tyr	Val	Ser	Pro	Glu	Phe	Trp	Leu	Leu	Thr	Lys
				260					265					270
Gln	Arg	Asn	Pro	Met	Asp	Tyr	Pro	Val	Glu	Asp	Ala	Phe	Cys	Lys
				275					280					285
Pro	Gln	Leu	Val	Lys	Lys	Ser	Tyr	Gly	Val	Glu	Asn	Arg	Ala	Tyr
				290					295					300
Ser	Gln	Glu	Glu	Ile	Thr	Gln	Gly	Phe	Glu	Glu	Thr	Gly	Asp	Thr
				305					310					315
Leu	Tyr	Ala	Pro	Tyr	Ser	Thr	His	Phe	Gln	Leu	Gln	Asn	Gln	Pro
				320					325					330
Pro	Gln	Lys	Glu	Phe	Ser	Ile	Pro	Arg	Ala	His	Ala	Trp	Pro	Ser
				335					340					345
Pro	Tyr	Lys	Asp	Tyr	Glu	Val	Lys	Lys	Glu	Gly	Ser			
				350					355					

&lt;210&gt; 22

&lt;211&gt; 1069

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2859482CD1

&lt;400&gt; 22

Met	Asp	Asp	Lys	Ala	Ser	Val	Gly	Lys	Ile	Ser	Val	Ser	Ser	Asp
1				5					10					15
Ser	Val	Ser	Thr	Leu	Asn	Ser	Glu	Asp	Phe	Val	Leu	Val	Ser	Arg
				20					25					30
Gln	Gly	Asp	Glu	Thr	Pro	Ser	Thr	Asn	Asn	Gly	Ser	Asp	Asp	Glu
				35					40					45
Lys	Thr	Gly	Leu	Lys	Ile	Val	Gly	Asn	Gly	Ser	Glu	Gln	Gln	Leu
				50					55					60
Gln	Lys	Glu	Leu	Ala	Asp	Val	Leu	Met	Asp	Pro	Pro	Met	Asp	Asp
				65					70					75
Gln	Pro	Gly	Glu	Lys	Glu	Leu	Val	Lys	Arg	Ser	Gln	Leu	Asp	Gly
				80					85					90
Glu	Gly	Asp	Gly	Pro	Leu	Ser	Asn	Gln	Leu	Ser	Ala	Ser	Ser	Thr



	95		100		105
Ile Asn Pro Val	Pro Leu Val Gly Leu	Gln Lys Pro Glu Met	Ser		
	110		115		120
Leu Pro Val Lys	Pro Gly Gln Gly Asp	Ser Glu Ala Ser Ser	Pro		
	125		130		135
Phe Thr Pro Val	Ala Asp Glu Asp Ser	Val Val Phe Ser Lys	Leu		
	140		145		150
Thr Tyr Leu Gly	Cys Ala Ser Val Asn	Ala Pro Arg Ser Glu	Val		
	155		160		165
Glu Ala Leu Arg	Met Met Ser Ile Leu	Arg Ser Gln Cys Gln	Ile		
	170		175		180
Ser Leu Asp Val	Thr Leu Ser Val Pro	Asn Val Ser Glu Gly	Ile		
	185		190		195
Val Arg Leu Leu	Asp Pro Gln Thr Asn	Thr Glu Ile Ala Asn	Tyr		
	200		205		210
Pro Ile Tyr Lys	Ile Leu Phe Cys Val	Arg Gly His Asp Gly	Thr		
	215		220		225
Pro Glu Ser Asp	Cys Phe Ala Phe Thr	Glu Ser His Tyr Asn	Ala		
	230		235		240
Glu Leu Phe Arg	Ile His Val Phe Arg	Cys Glu Ile Gln Glu	Ala		
	245		250		255
Val Ser Arg Ile	Leu Tyr Ser Phe Ala	Thr Ala Phe Arg Arg	Ser		
	260		265		270
Ala Lys Gln Thr	Pro Leu Ser Ala Thr	Ala Ala Pro Gln Thr	Pro		
	275		280		285
Asp Ser Asp Ile	Phe Thr Phe Ser Val	Ser Leu Glu Ile Lys	Glu		
	290		295		300
Asp Asp Gly Lys	Gly Tyr Phe Ser Ala	Val Pro Lys Asp Lys	Asp		
	305		310		315
Arg Gln Cys Phe	Lys Leu Arg Gln Gly	Ile Asp Lys Lys Ile	Val		
	320		325		330
Ile Tyr Val Gln	Gln Thr Thr Asn Lys	Glu Leu Ala Ile Glu	Arg		
	335		340		345
Cys Phe Gly Leu	Leu Leu Ser Pro Gly	Lys Asp Val Arg Asn	Ser		
	350		355		360
Asp Met His Leu	Leu Asp Leu Glu Ser	Met Gly Lys Ser Ser	Asp		
	365		370		375
Gly Lys Ser Tyr	Val Ile Thr Gly Ser	Trp Asn Pro Lys Ser	Pro		
	380		385		390
His Phe Gln Val	Val Asn Glu Glu Thr	Pro Lys Asp Lys Val	Leu		
	395		400		405
Phe Met Thr Thr	Ala Val Asp Leu Val	Ile Thr Glu Val Gln	Glu		
	410		415		420
Pro Val Arg Phe	Leu Leu Glu Thr Lys	Val Arg Val Cys Ser	Pro		
	425		430		435
Asn Glu Arg Leu	Phe Trp Pro Phe Ser	Lys Arg Ser Thr Thr	Glu		
	440		445		450
Asn Phe Phe Leu	Lys Leu Lys Gln Ile	Lys Gln Arg Glu Arg	Lys		
	455		460		465
Asn Asn Thr Asp	Thr Leu Tyr Glu Val	Val Cys Leu Glu Ser	Glu		
	470		475		480
Ser Glu Arg Glu	Arg Arg Lys Thr Thr	Ala Ser Pro Ser Val	Arg		
	485		490		495
Leu Pro Gln Ser	Gly Ser Gln Ser Ser	Val Ile Pro Ser Pro	Pro		
	500		505		510
Glu Asp Asp Glu	Glu Glu Asp Asn Asp	Glu Pro Leu Leu Ser	Gly		
	515		520		525
Ser Gly Asp Val	Ser Lys Glu Cys Ala	Glu Lys Ile Leu Glu	Thr		
	530		535		540
Trp Gly Glu Leu	Leu Ser Lys Trp His	Leu Asn Leu Asn Val	Arg		
	545		550		555
Pro Lys Gln Leu	Ser Ser Leu Val Arg	Asn Gly Val Pro Glu	Ala		
	560		565		570
Leu Arg Gly Glu	Val Trp Gln Leu Leu	Ala Gly Cys His Asn	Asn		

	575		580		585
Asp His Leu Val	Glu Lys Tyr Arg Ile	Leu Ile Thr Lys Glu Ser			
	590		595		600
Pro Gln Asp Ser	Ala Ile Thr Arg Asp	Ile Asn Arg Thr Phe Pro			
	605		610		615
Ala His Asp Tyr	Phe Lys Asp Thr Gly	Gly Asp Gly Gln Asp Ser			
	620		625		630
Leu Tyr Lys Ile	Cys Lys Ala Tyr Ser	Val Tyr Asp Glu Glu Ile			
	635		640		645
Gly Tyr Cys Gln	Gly Gln Ser Phe Leu	Ala Ala Val Leu Leu Leu			
	650		655		660
His Met Pro Glu	Glu Gln Ala Phe Ser	Val Leu Val Lys Ile Met			
	665		670		675
Phe Asp Tyr Gly	Leu Arg Glu Leu Phe	Lys Gln Asn Phe Glu Asp			
	680		685		690
Leu His Cys Lys	Phe Tyr Gln Leu Glu	Arg Leu Met Gln Glu Tyr			
	695		700		705
Ile Pro Asp Leu	Tyr Asn His Phe Leu	Asp Ile Ser Leu Glu Ala			
	710		715		720
His Met Tyr Ala	Ser Gln Trp Phe Leu	Thr Leu Phe Thr Ala Lys			
	725		730		735
Phe Pro Leu Tyr	Met Val Phe His Ile	Ile Asp Leu Leu Leu Cys			
	740		745		750
Glu Gly Ile Ser	Val Ile Phe Asn Val	Ala Leu Gly Leu Leu Lys			
	755		760		765
Thr Ser Lys Asp	Asp Leu Leu Leu Thr	Asp Phe Glu Gly Ala Leu			
	770		775		780
Lys Phe Phe Arg	Val Gln Leu Pro Lys	Arg Tyr Arg Ser Glu Glu			
	785		790		795
Asn Ala Lys Lys	Leu Met Glu Leu Ala	Cys Asn Met Lys Ile Ser			
	800		805		810
Gln Lys Lys Leu	Lys Lys Tyr Glu Lys	Glu Tyr His Thr Met Arg			
	815		820		825
Glu Gln Gln Ala	Gln Gln Glu Asp Pro	Ile Glu Arg Phe Glu Arg			
	830		835		840
Glu Asn Arg Arg	Leu Gln Glu Ala Asn	Met Arg Leu Glu Gln Glu			
	845		850		855
Asn Asp Asp Leu	Ala His Glu Leu Val	Thr Ser Lys Ile Ala Leu			
	860		865		870
Arg Lys Asp Leu	Asp Asn Ala Glu Glu	Lys Ala Asp Ala Leu Asn			
	875		880		885
Lys Glu Leu Leu	Met Thr Lys Gln Lys	Leu Ile Asp Ala Glu Glu			
	890		895		900
Glu Lys Arg Arg	Leu Glu Glu Glu Ser	Ala Gln Leu Lys Glu Met			
	905		910		915
Cys Arg Arg Glu	Leu Asp Lys Ala Glu	Ser Glu Ile Lys Lys Asn			
	920		925		930
Ser Ser Ile Ile	Gly Asp Tyr Lys Gln	Ile Cys Ser Gln Leu Ser			
	935		940		945
Glu Arg Leu Glu	Lys Gln Gln Thr Ala	Asn Lys Val Glu Ile Glu			
	950		955		960
Lys Ile Arg Gln	Lys Val Asp Asp Cys	Glu Arg Cys Arg Glu Phe			
	965		970		975
Phe Asn Lys Glu	Gly Arg Val Lys Gly	Ile Ser Ser Thr Lys Glu			
	980		985		990
Val Leu Asp Glu	Asp Thr Asp Glu Glu	Lys Glu Thr Leu Lys Asn			
	995		1000		1005
Gln Leu Arg Glu	Met Glu Leu Glu Leu	Ala Gln Thr Lys Leu Gln			
	1010		1015		1020
Leu Val Glu Ala	Glu Cys Lys Ile Gln	Asp Leu Glu His His Leu			
	1025		1030		1035
Gly Leu Ala Leu	Asn Glu Val Gln Ala	Lys Lys Thr Trp Phe			
	1040		1045		1050
Asn Arg Thr Leu	Ser Ser Ile Lys Thr	Ala Thr Gly Val Gln Gly			

Lys Glu Thr Cys 1055

1060

1065

<210> 23  
 <211> 1995  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <221> misc\_feature  
 <223> Incyte ID No: 209171CB1

<400> 23  
 ggaccgtctt cccagtcacca tcaagagaaa accacagact ctgggctcac tgaaggcata 60  
 tggcagctgg tacctccatc actgtttaaa ggctcacata tcagtcaggg aaacgagggt 120  
 gaggaagag aggagccttg ggaccacact gaaaaaactg aagaggagcc ggtctctggc 180  
 agctcaggaa gctgggacca gtcaagccag ccagtgtttg agaatgtgaa cggttaaattct 240  
 tttagacagat gtactggcca ctctggctgag cacacacagt gtgggaagcc acaggaaagt 300  
 actgggaggg gttctgcttt tctcaaagct gtccagggtg gcggggacac atctaggcac 360  
 tgtctacctt ccctagcaga tgccaaaggt ctccaggaca ctggggggcac tgtgaactat 420  
 ttctggggta ttccattctg ccctgatgga gtagacccta accagtatac caaggtcatt 480  
 ctctgccagt tggaggttta tcaaaagagc ctgaaaatgg ctccagaggca gctccttaat 540  
 aaaaaaggtt ttgggggaacc agtggttacct agacctcctt ctctgatcca gaatgaatgt 600  
 ggccaaggag agcaggctag tgagaaaaat gaatgcatct cagaagatat gggagatgaa 660  
 gacaaagagg agaggcagga gtctagggca tctgactggc actcaaaaac caaggatttc 720  
 caggaaagct caattaaaag cttgaaagag aaacttttgt tggagggaaga accaacaacc 780  
 agtcatggtc agtcttccca agggattgtt gaagaaactt ctgaagaggg aaactctgta 840  
 cctgcttcac aaagtgttgc tgctttgacc agtaagagaa gcttagtcct tatgccagag 900  
 agttctgcag aagaaatcac tgtttgcct gagaccagc taagttcctc tgaaactttt 960  
 gaccttgaag gagaagtctc tccaggtagc agagatatct tggatggagt cagaataata 1020  
 atggcagata aggaggttgg taacaaggaa gatgctgaga aggaagtagc tatttctacc 1080  
 ttctcatcca gtaaccaggt atcctgcccg ctatgtgacc aatgctttcc acccacaag 1140  
 attgaacgac atgccatgta ctgcaatggt ctgatggagg aagatacagt attgactcgg 1200  
 agacaaaaag agggcaagac caagagtgc agtgggacag ctgcccagac ttctctagac 1260  
 attgacaaga atgagaagtg ttacctctgt aaatccctgg tcccathtag agagtatcag 1320  
 tgtcatgtgg actcctgtct ccagcttgca aaggctgacc aaggagatgg acctgaaggg 1380  
 agtggaagag catgttcaac tgtggagggg aagtggcagc agaggctgaa gaacccaaag 1440  
 gaaaaaggcc acagtgaagg ccgactcctt agtttcttgg aacagtctga gcacaagact 1500  
 tcagatgcag acatcaagtc ttcagaaaca ggagccttca gggtgccctc accagggatg 1560  
 gaagaggcag gctgcagcag agagatgcag agttctttca cagctcgtga cttaaatagaa 1620  
 tctcccgta agtcttttgt ttccatttca gaagccacag attgcttagt ggactttaaa 1680  
 aagcaagtta ctgtccagcc aggtagtcgg acacggacca aagctggcag aggaagaagg 1740  
 agaaaattct gaatttctag ggtccaaaag ttgacaaaac cattagtagg aggggtgggc 1800  
 catgttcatt aagccatagt ggtccctagt tcattgttga gcaagtttta gccctgcagt 1860  
 tttcaccacc agcacctacc cagcattctg gttttatgt tttttatgat ctatgcagac 1920  
 aactgtgtat tctgttttat aacagtttgt ttgaatttac ttacagttaa aaaattttaa 1980  
 tataaaaaaa aaaaa 1995

<210> 24  
 <211> 2051  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 945430CB1

<400> 24  
 tcgaccacgc gtccgtgcgt gttcactgtt tagtagtact caaaactgcc agtgtgagag 60  
 gatttggaat tcaactggatc tgctcaatac aaaaatgttt ttctcagagt ttctccattt 120  
 tatcaggagt cattctgcca ctgcagtgga tttccttctt gtgatgggtg accggctccc 180  
 agttttcaaa aggtacatgg gaaatactcc tcagaaaaaa gccgtctttg ggcagtgtcg 240  
 gggctctgcca tgtgttgac cgctgctgac cacagtggaa gaggctccac ggggcatctc 300  
 tgctcagatc tggggacatt ttcctaagtg gctcaatggc tctctacttc gaattggacc 360

```

tgggaaattc gagtttggga aggataagta caatcattgg tttgatggga tggcgctgct 420
tcaccagttc agaatggcaa agggcacagt gacatacagg agcaagtttc tacagagtga 480
tacatataag gccaacagtg ctaaaaaccg aa:tgatgc tcagaatttg gcacactggc 540
tctcccgat ccatgcaaga atgtttttga acgtttcatg tccaggtttg agctgcctgg 600
taaagctgca gccatgactg acaatactaa tgtcaactat gtgcggtaca aggggtgatta 660
ctacctctgc actgagacca actttatgaa taaagtggac attgaaactc tggaaaaaac 720
agaaaaggta gattggagca aatttatgtc tgtgaatgga gcaactgcac atcctcatta 780
tgaccggat ggaacagcat acaatatggg gaactccttt gggccatatg gtttctccta 840
taaggttatt cgggttcctc cagagaagggt ggaccttggg gagacaatcc atggagtcca 900
ggtgatattg tctattgctt ctacagagaa agggaaacct tcttactacc atagctttgg 960
aatgacaagg aactatataa ttttcattga acaacctcta aagatgaacc tgtggaaaat 1020
tgccacttct aaaattcggg gaaaggcctt ttcagatggg ataagctggg aaccccagtg 1080
taatacgcg tttcatgtgg tggaaaaacg cactggacag ctccttccag ggagatacta 1140
cagcaaacct tttgttacat ttcatacaat caatgccttt gaggaccagg gctgtgttat 1200
aattgatttg tgctctcaag ataatggaag aacctagaa gtttaccagt tacagaatct 1260
caggaaggct ggggaagggt ttgatcagggt ccataattca gcagccaaat ctttccctcg 1320
aaggtttgtt ttgcctttta atgtcagttt gaatgccctt gagggagaca acctgagtcc 1380
attgtcctat acttcagcca gtgctgtgaa acaggctgat ggaacgatct ggtgctctca 1440
tgaaaatcta catcaggagg acctagaaaa ggaaggaggc attgaatttc ctcatgctta 1500
ctatgatcga ttcagtggca aaaagtatca tttcttttat ggctgtggct ttcggcattt 1560
agtgggggat tctctgatca aggttgatgt ggtgaataag aactgaagg tttggagaga 1620
agatggcttt tatccctcag aacctgtttt tgttccagca ccaggaacca atgaagaaga 1680
tggtggggtt attctttctg tggatgacac tcccaaccag aatgaaagca attttctct 1740
agttttggat gccagaactt ttgaagagct gggccgagca gaggtacctg tgcagatgcc 1800
ttatgggttc catggtacct tcatacccat ctgatgggac aaccacaagg tctggaaact 1860
aggtttaaaa taagtgtgca cttggacata aagactggag aaataaacac tgaggactcc 1920
aaaagggggg caaggaggaa gagggggcagg ggttaaaaag ctacctattg aatactatgt 1980
tccttatttg ggtgatgggt tcattagaag tccaaacctc agcagcacac aatatactca 2040
tgtaacaagg g
2051

```

&lt;210&gt; 25

&lt;211&gt; 2067

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1305513CB1

&lt;400&gt; 25

```

ggcaaaaagc atgcagaaaa agaagcagac gttttacatt gggaattaat gaaagcgtgt 60
ctgctagttt tgggtaggag aactgggaag ttgttgctta aaattttata tcacctccac 120
aaacaaaact cttcggaagt ggtaaaaata gaaatgcat gattctagag gcattcctaa 180
gcacccacgt gtcgggcttt gtggtgtctg tggatcatc cgaccgtttg gactgggttag 240
ggcttactga gagctccatt tctggaaagc cttacaagac tgaggaaatag cagactgcga 300
atcacccgga acggttcctt tgcagcacag aagcaatctc tctccccatc ttcgcatatt 360
ctgatggcaa aacaagtggg agaaaagagg aagcatgact gcagatcaga tcagttctct 420
ttgtggatta tattttcagt aaaatgtatg gatctatctt ttccttgctt ttatatctag 480
atcatgagac ttgactgagg ctgtatcctt atcctccatc catctatggc gaactatagc 540
catgcagctg acaacatttt gcaaaatctc tcgctcttaa cagcctttct gaaactgact 600
tccttgggtt tcataatagg agtcagcgtg gtgggcaacc tctgatctc cattttgcta 660
gtgaaagata agaccttgca tagagcacct tactacttcc tggttggatct ttgctgttca 720
gatatactca gatctgcaat ttgtttccca tttgtgttca actctgtcaa aaatggctct 780
acctggactt atgggactct gacttgcaaa gtgattgcct tcttgggggt tttgtcctgt 840
ttccacactg ccttctatgt cttctgcac agtgtcacca gatacttagc tatcgcccat 900
caccgcttct atacaaagag gctgacctt ttgacgtgtc tggctgtgat ctgtatggg 960
tggactctgt ctgtggccat ggcatttccc cgggttttag acgtgggcac ttactcattc 1020
attagggaga aagatcaatg caccttccaa caccgctcct tcagggctaa tgattcctta 1080
ggatttatgc tgcttcttgc tctcatctc ctagccacac agcttgtcta cctcaagctg 1140
atatttttcg tccacgatcg aagaaaaatg aagccagtc agttttagc agcagtcagc 1200
cagaactgga cttttcatgg tcctggagcc agtggccagg cagctgcaa ttggctagca 1260
ggatttgga ggggtcccac accaccacac ttgctgggca tcaggcaaaa tgcaaacacc 1320
acaggcagaa gaaggctatt ggtcttagac gagttcaaaa tggagaaaag aatcagcaga 1380
atgttctata taatgacttt tctgtttcta accttgtggg gccctacctt ggtggcctgt 1440

```

```

tattggagag tttttgcaag agggcctgta gtaccagggg gattttctaac agctgctgtc 1500
tggatgagtt ttgcccgaagc aggaatcaat ccttttgtct gcattttctc aaacagggag 1560
ctgaggcgct gtttcagcac aacccttctt tactgcagaa aatccagggtt accaagggaa 1620
ccttactgtg ttatatgagg gagcatctgt aaatcttttag ccttgtgaaa actaaccttc 1680
tctgctgagc aattgtggcc catagccata ttttgagaag aaattcaaga atggaatcag 1740
cagttttaag gatttgggca acattctgca gtctttgcaa tagttcacct ataatcctat 1800
tttaaatctc agagtgatcc tgctgactgc cagcaaagggt ttgtaattaa gaagggactg 1860
aaccactgcc ctaagtttct ttatgtggtc aaaaactaga taatgaaagt agcagggtgct 1920
aagtatcagt gctaaatgct ctgtatgtca ctacatatga aaaaacatca aaaaacaatt 1980
agcattggac atcttaataa attaagttga catgaggtaa atgtgttgat aaaaactaat 2040
tttagaagtt tgaagacttt aaaacag 2067

```

<210> 26  
 <211> 1165  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1876283CB1

```

<400> 26
cttcttccgg gtgggggcccc gggccgaggc gatggcgccc tgggcgctcc tcagccctgg 60
ggtcctggtg cggaccgggc acaccgtgct gacctgggga atcacgctgg tgctcttctc 120
gcacgatacc gagctgcggc aatgggagga gcagggggag ctgctcctgc cctcacctt 180
cctgctcctg gtgctgggct cctgctgct ctacctgct gtgtcactca tggaccctgg 240
ctacgtgaat gtgcagcccc agcctcagga ggagctcaaa gaggagcaga cagccatgg 300
tcctccagcc atccctcttc ggcgctgcag atactgctg gtgctgcagc cctgagggc 360
tcggcactgc cgtgagtgc gccgttgct cgcgcgtac gaccaccact gcccctggat 420
ggagaactgt gtgggagagc gcaaccacc actctttgtg gtctacctgg cgctgcagct 480
ggtggtgctt ctgtggggcc tgtacctggc atggctcagg ctccggttct tccagccctg 540
gggtctgtgg ttgcgggtcca gggggtcct gttegccacc ttctgctgc tgctcctctt 600
ctcgttggtg gccagcctgc tctcgtctc gcacctctac ctggtggcca gcaacaccac 660
cacctgggaa ttcactctct cacaccgat cgcctatctc cgccagcgcc ccagcaaccc 720
cttcgaccga ggccctgacc gcaacctggc ccacttcttc tgtggatggc cctcagggtc 780
ctgggagacc ctctgggctg aggaggagga agaggcgagc agcccagctg tttaggggtg 840
ctggaggccg ggctaccgct ttgtgcctga aaaccacggg gcctgtcccc agctgggggtg 900
agcgtcaga gggcctgggg ccctcactcc tgcccacgcc tcccagacct cagaacggag 960
cttcaagtca gacagatccc tgccctgggtg ggcagttctg ccttccaagg aagaagggga 1020
agaaaaggac ctgtgggtgg ctcaggccca agcagacccc gggctccacc ccagccccgc 1080
ccaggctgct gccagtgcac acttttacia atttaatata aagcaagtc agtcttaaaa 1140
agacaaacca taaaaataaa aaaaa 1165

```

<210> 27  
 <211> 3523  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2470285CB1

<220>  
 <221> unsure  
 <222> 3393, 3419, 3435, 3443  
 <223> a, t, c, g, or other

```

<400> 27
ggggccgcag caatgccggg cccgctaggg ctgctctgct tctcgcctt ggggctgctc 60
ggctcggccg ggcccagcgg cgcggcgccg cctctctgct cggcgcctt cagctgcgac 120
ggcgaccgtc ggggtgactg ctccgggaag gggctgacgg ccgtgcccg ggggctcagc 180
gccttcaccc aagcgttgga tatcagtatg aacaacatta ctcagttgcc agaagatgca 240
tttaagaact ttcttttctt agaagagcta caattggcgg gcaacgacct ttcttttatc 300
caccctaaag ccttgtctgg gttgaaagaa ctcaaagttc taacgctcca gaataatcag 360

```

```

ttgaaaacag taccacagtga agccattcga gggctgagtg ctttgcagtc tttgcgttta 420
gatgccaacc atattacctc agtccccgag gacagttttg aaggacttgt tcagttacgg 480
catctgtggc tggatgacaa cagcttgacg gaggtgcctg tgcacccctc cagcaatctg 540
cccaccctac aggcgctgac cctggctctc aacaagatct caagtatccc tgactttgca 600
tttaccaacc tttcaagcct ggtagtctctg catcttcata acaataaaat tagaagcctg 660
agtcaacact gttttgatgg actagataac ctggagacct tagacttgaa ttataataac 720
ttgggggaat ttcttcaggc tattaagacc ctctctagcc ttaaagagct aggatttcat 780
agtaattcta tttctgttat ccctgatgga gcatttgatg gtaatccact cttagaact 840
atacatttgt atgataatcc tctgtctttt gtggggaact cagcatttca caatttatct 900
gatcttcatt ccctagtcac tctgtgtgca agcatggtgc agcagttccc caatcttaca 960
ggaactgtcc acctggaag tctgactttg acaggtacaa agataagcag catacctaata 1020
aatttgtgcc aagaacaaaa gatgcttagg actttggact tgtcttaca taatataaga 1080
gaccttccaa gttttaatgg ttgccatgct ctggaagaaa tttctttaca gcgtaataca 1140
atctaccaa taaaggaagg cacctttcaa ggctgatata ctctaaggat tctagatctg 1200
agtagaaacc tgatacatga aattcacagt agagcttttg ccacacttgg gccataact 1260
aacctagatg taagtttcaa tgaattaaat tcctttccta cggaaggcct gaatgggcta 1320
aatcaactga aacttgtggg caacttcaag ctgaaagaag ccttagcagc aaaagacttt 1380
gttaacctca ggtctttatc agtaccatat gcttatcagt gctgtgcatt ttggggttgt 1440
gactcttatg caaatttaaa cacagaagat aacagcctcc aggaccacag tgtggcacag 1500
gagaaaggta ctgctgatgc agcaaatgtc acaagcactc ttgaaaatga agaacatagt 1560
caaataatta tccattgtac accttcaaca ggtgctttta agccctgtga atatttactg 1620
ggaagctgga tgattcgtct tactgtgtgg ttcatcttct tggttgcatt atttttcaac 1680
ctgcttggtta ttttaacaac atttgcattc ttacatcac tgccttcgtc caaattgttt 1740
ataggcttga tttctgtgtc taacttatct atgggaatct atactggcat cctaactttt 1800
cttgatgctg tgtcctgggg cagattcgct gaatttggca tttgggtggg aactggcagt 1860
ggctgcaaag tagctgggtt tcttgcagtt ttctcctcag aaagtgccat atttttatta 1920
atgctagcaa ctgtcgaag aagcttatct gcaaaagata taatgaaaaa tgggaagagc 1980
aatcatctca aacagttccg ggttgctgcc cttttggctt tcctaggtgc tacagtagca 2040
ggctgttttc cctttttcca tagaggggaa tacttgcac caccctttg tttgccattt 2100
cctacaggtg aaacgccatc attaggattc actgtaacgt tagtgctatt aaactcacta 2160
gcatttttat taatggccgt tatctacact aagctatact gcaacttggg aaaagaggac 2220
ctctcagaaa actcacaatc tagcatgatt aagcatgtcg cttggctaata cttcaccaat 2280
tgcattcttt tctgccctgt ggcgtttttt tcatattgcac cattgatcac tgcaatctct 2340
atcagccccc aaataatgaa gtctgttaact ctgatatttt ttccattgcc tgcttgccctg 2400
aatccagtc tgtatgtttt ctccaaccca aagttttaaag aagactggaa gttactgaag 2460
cgacgtgta ccaagaaaag tggatcagtt tcagtttcca tcagtagcca aggtgggtgt 2520
ctggaacagg atttctacta cgactgtggc atgtactcac atttgcaggg caacctgact 2580
gtttgcgact gctgcgaatc gtttctttta acaaagccag tatcatgcaa acacttgata 2640
aaatcacaca gctgtcctgc attggcagtg gcttcttgcc aaagacctga gggctactgg 2700
tccgactgtg gcacacagtc ggccactct gattatgcag atgaagaaga ttcctttgtc 2760
tcagacagtt ctgaccaggt gcaggcctgt ggaagagcct gcttctacca gagtagagga 2820
ttccctttgg tgcctatgc ttacaatcta ccaagagtta aagactgaac tactgtgtgt 2880
gtaaccgttt ccccgtcaa ccccgctcag tgtttataga gtgaacccta ttctcatctt 2940
tcatctggga agcacttctg taatcactgc ctggtgtcac ttagaagaag gagaggtggc 3000
agttttatttc tcaaaccagt cattttcaaa gaacaggtgc cttaaattata aattggtgaa 3060
aaatgcaatg tccaagcaat gtatgatctg tttgaaacaa atatatgact tgaaaaggat 3120
cttaggtgta gtagagcaat ataattgttag ttttttctga tccataagaa gcaaatttat 3180
acctatttgt gtattaagca caagataaag aacagctgtt aatatttttt aaaaatctat 3240
tttaaaatgt gattttctat aactgaagaa aatatcttgc taattttacc taatgtttca 3300
tcttaactct caggacaact tactgcaggg ccaaaaaagg gactgtccca gctagaactg 3360
tgagagtata cataggcatt actttattat gtnttcaact gccatccttg acataagang 3420
actataaatt ttgtntaagc canttataaa tctaaacctg gtatagtgtt taaaccatat 3480
tacagctgtt aggttaaaaa atagctggac attggtttca gtt 3523

```

&lt;210&gt; 28

&lt;211&gt; 2179

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: P2925789CB1

&lt;400&gt; 28

```

gtcctggcga gggcgctggc cgagaggtgc tcggcttgta gcaggtcccg cactccagcc 60
tctcgctgcc agggtttgtc ctctgcttgc cctgggctga ggtgtccatg acggagtcac 120
ccaaggagga aaaaatctgt tccgggtgag cccaggccgc cccggatatg cgatggctga 180
ggagcagaca ccagggacca cactgagggtt gggtttcaga ccaagacact ggattctcct 240
agttaagata aagagctttg ggtgcctgac agtgaaaatg gtgtaatctg cgttaacagt 300
tcacagcttg aaggcatgac aattaaagaa cacacatgga cttgtggcac atggaaatgt 360
gcgcacagaa aaaggaaatc tataattctt ttaaagtagg aaggcattct tccttgccaa 420
aatgggtaca ttctgttcgg ttatcaagtt tgaaaatcta caagaattaa agagactgtg 480
tcaactgggt cccatcatag cccttgggtg tacaactgga ggaagtgtga atttcatcat 600
tgactctgtg ttgtggtatt ggcccttaca tacaactgga aatgccatgt ttgtcgggtc 660
gttgataaat tggactgtca tgattcttta taattacttc aatgccatgt ttgtcgggtc 660
gggctttgtc cctctggggg ggaaccgga aatttctcag gataccatgt atctccagta 720
ttgtaaagtc tgccaagcat acaaggcacc acgttcacat cactgcagaa agtgaacag 780
atgtgtgatg aagatggacc atcactgtcc ttggatcaac aactgttgtg gttacacaaa 840
tcatgcttcg ttcacactgt ttctcctttt agcaccactg ggtgtatcc atgctgcttt 900
catttttgtg atgactatgt acacacagct tttatcctgg ctctcctttg ggtggaacac 960
agtgaagatg gacatgagtg cagcccgag agatcctctt ccaattgttc catttggatt 1020
agctgcattt gctaccacct tgtttgcctt gggattagct ttaggaacaa ccatagctgt 1080
tgggatgttg ttttttatcc agatgaaaat aatttctcaga aacaaaactt ctattgagtc 1140
atggattgaa gagaaggcta aagatcgaat tcagtattat caactagatg aagtctttgt 1200
ttttccatat gatatgggaa gtagatggag gaactttaaa caggtattta cgtggtcagg 1260
ggtccctgaa ggagatggac ttgagtggcc agtaagagaa ggctgtcacc aatacagctt 1320
aacaatagaa cagttgaaac aaaaagcaga taagagagtc agaagtgttc gctataaagt 1380
aatagaagat tatagtgggt cctgctgccc tctgaataaa ggaatcaaaa ccttcttcac 1440
aagtccttgc accgaagagc ctgcaataca gctgcaaaaa ggggaattca ttttagccac 1500
aagaggttta cgatactggt tatatggaga caaaattctt gatgattcct ttatagaagg 1560
tgtttcaaga ataaggggtt ggttccctag aaaatgtgtg gaaaagtgtc cctgtgatgc 1620
tgaaacagat caagcccag agggggagaa gaaaaataga tagctgctgt taaaacaaa 1680
ttatccttta agtctgctta attacttgaa aattgtacat attactaaag aattatgcaa 1740
tgagcctact ctggtttaaga tgttcttttc ctcaaagggt ccctagtgc atgattttaa 1800
tatttttatt accattttga aatggagaag ccattctgca tatgcctttg aattcctgcc 1860
cctctttacc acctcttcct cccctcaaa ggaaaaacat ttcacccaag taagttaacg 1920
gcattttctg taggattttc ttatgcactg cacactctgg acctcacctg cagatacagt 1980
tcccccttg ccaggagcat ctgcatgtgg tacttctctt tccctcagt tgatatttct 2040
tatatgatat tctagatact atagaactca atttgtcaga ttcagtataa cctcagattt 2100
tgttacctgt cttttaaaaa tgcagatttt gtcaaatcaa ataaagatca atggatgttg 2160
ggtataaaaa aaaaaaaaaa

```

&lt;210&gt; 29

&lt;211&gt; 645

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3099990CB1

&lt;400&gt; 29

```

cggccaaaga tctcagcctt cctgcccgcc cggcagctct ggaagtgggtc ggggaatccc 60
acacagcggc gtggcatgaa ggggaaggcc cggaaagctgt tctacaaggc catcgtgcgg 120
ggcgaggaga ccctgcgtgt cggggactgt gccgtcttcc tgtcagctgg gcggcccaac 180
ctccccatac tcggccgcat cgagagcatg tgggagtcgt ggggcagcaa catggtgggtc 240
aaggtcaagt ggttctacca ccctgaggag accaagctgg gcaagaggca gtgcgacggc 300
aagaatgcgc tgtaccagtc ctgccacgag gatgagaacg acgtgcagac catctccac 360
aagtgccagg tcgtggcgcg cgagcagtat gagcagatgg cccggagccg caagtgccag 420
gaccggcagg acctctacta cctggcgggc acctacgacc ccaccaccgg gcgcctgggtg 480
acggctgatg gcgtgcccat cctatgctga gccgcccacc gcagatgcct cccacgtgcg 540
ccaggaccc tgtgtgcgga cctggcgctg gccaaagccac cgggcaggag gcagccccgg 600
cctcccaagg gcgcatctga gcaaatatgc aaaagccac agggc 645

```

&lt;210&gt; 30

&lt;211&gt; 627

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 103561CB1

&lt;400&gt; 30

```
cggactcaga agggacttcc ctgctcgggt ggctttcggg ttctctgctc acctccggat 60
aaatcacggg gtctcccgcg ccgctcatgg cgcctcccggt ccgtctcgag cgtccctttc 120
cttcccggcg ctttctctggg ttgcttctgg cggccctggg gttgctgctg tctctcttct 180
ccgatcaatg caatgtcccg gaatggcttc catttgccag gcctaccaac ctaactgatg 240
actttgagtt tcccattggg acatatctga actatgaatg ccgccctggg tattccggaa 300
gaccgttttc tatcatctgc ctaaaaaact cagtctggac aagtgctaag gacaagtgca 360
aacgtaaatc atgtcgtaat cctccagatc ctgtgaatgg catggcacat gtgatcaaaag 420
acatccagtt cggatcccaa attaaatatt cttgtcctaa aggataccga ctcatgggtt 480
cctcgtctgc cacatgcata atctcaggca acaactgtcat ttgggataat aaaacacctg 540
tttgtgacag tgagttgaaa tatgcattcc tatttctttt accgatacat tctaattttt 600
ctctggaata ataaaaatct taaccga 627
```

&lt;210&gt; 31

&lt;211&gt; 1858

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 288709CB1

&lt;400&gt; 31

```
ctggatgagc tgctggaggc cctggggacc tggagaccct gccccgctta cagcagcact 60
gtgtggccag tcacctgctg gatggcctag aggatgtcct cagaggcctg agcaagaacc 120
tttccaatgg gctgttgaaac ttcagttatc ctgcaggcac agaattgtcc ctggaggtgc 180
agaagcaagt agacaggagt gtcaccttga gacagaatca ggcagtgatg cagctcgact 240
ggaatcaggc acagaaatct ggtgacccag gcccttctgt ggtgggcctt gtctccattc 300
cagggatggg caagttgctg gctgaggccc ctctggctct ggaacctgag aagcagatgc 360
ttctgcatga gacacaccag ggcttgctgc aggacggctc ccccatcctg ctctcagatg 420
tgatctctgc ctttctgagc aacaacgaca cccaaaacct cagctcccca gttaccttca 480
ccttctccca ccgttcagtg atcccgagac agaagggtgt ctgtgtcttc tgggagcatg 540
gccagaatgg atgtggtcac tgggccacca cagggtgcag cacaataggc accagagaca 600
ccagcaccat ctgccgttgc acccacctga gcagctttgc cgtctcatg gccactacg 660
atgtgcagga ggaggatccc gtgctgactg tcatcaccta catggggctg agcgtctctc 720
tgctgtgcct cctcctggcg gccctcactt ttctcctgtg taaagccatc cagaacacca 780
gcacctcact gcactctgag ctctcgtctt gcctcttctt ggcccacctc ctcttctctg 840
tggggattga tcgaactgaa cccaagggtg tgtgtccat catcgccggt gctttgcaact 900
atctctacct ggccgccttc acctggatgc tgctggaggg tgtgcacctc ttcctcactg 960
cacggaacct gacagtgggt aactactcaa gcatcaatag actcatgaag tggatcatgt 1020
tcccagtcgg ctatggcggt cccgctgtga ctgtggccat ttctgcagcc tctggcctc 1080
acctttatgg aactgctgat cgatgctggc tccacctgga ccagggatc atgtggagtt 1140
tccttggccc agtctgtgcc attttctctg cgaatttagt attgtttatc ttggctcttt 1200
ggattttgaa aagaaaactt tcctccctca atagtgaagt gtcaaccatc cagaacacaa 1260
ggatgctggc tttcaaagca acagctcagc tcttcactct gggctgcaca tgggtgtctgg 1320
gcttgctaca ggtgggtcca gctgcccagg tcatggccta cctcttcacc atcatcaaca 1380
gcctccaagg cttcttcata ttcttgggtt actgctctct cagccagcag gtccagaaac 1440
aatacaaaaa gtggtttaga gagatcgtaa aatcaaaatc tgagtctgag acatacacac 1500
tttcagacaa gatgggtcct gactcaaaac ccagtgaggg ggatgttttt ccaggacaag 1560
tgaagagaaa atattaaaac tagaatattc aactccatat ggaaaatcat atccatggat 1620
ctctttggca ttatgaagaa tgaagctaag gaaaagggaa ttcattaaac atatcatcct 1680
tggagaggaa gtaatcaacc tttacttccc aagctgtttg ttctccacaa taggctctca 1740
acaaatgtgt ggtaaattgc atttctcttc actatggtgt attcagtcaa tgcttgtccc 1800
tggaaaccca aagcatgacc actgcaaaata tttccttgac tttttgtaaa aaaaaaaaa 1858
```

&lt;210&gt; 32

&lt;211&gt; 539

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens



&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 959893CB1

&lt;400&gt; 32

```

tttttggctg agaaggctgg gtctacattt caggccacat ttggggagac gaatggagtc 60
attcctggga ggtgttttgc tgattttgtg gcttcaagtg gactgggtga agagccaaaa 120
gatagaacag aattccgagg ccctgaacat tcaggagggt aaaacggcca ccctgacctg 180
caactatata aactattccc cagcatactt acagtggtag cgacaagatc caggaagagg 240
ccctgttttc ttgctactca tacgtgaaaa tgagaaagaa aaaaggaaag aaagactgaa 300
ggtcaccttt gataccaccc ttaaacagag tttgtttcat atcacagcct cccagcctgc 360
agactcagct aactacctct gtgctctagg gggaagagga acaaactcac ctttgggaca 420
ggcactcagc tagaagtggg actcaatatc cagaaccctg accctgccgt gtacaagctg 480
agagactcta aatccagtga caagtctgtc tgcctattaa cggttttgat tctcaaaaa 539

```

&lt;210&gt; 33

&lt;211&gt; 1126

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1414179CB1

&lt;400&gt; 33

```

tgccatctta ggggcgcctg ggcgtacggg tttctcgttg gaggcggcct tcgtggcagc 60
tgtagacgcc gggaaaaggc ataaagtccg ttggccgaca cctttctttc ctccggcctc 120
ggtagaaccg ccagcccgcg tccgaaggcg gaggcgaggg gaactggccg cgtgaggggc 180
ctgaggcgag cgggttagagc gtctcccgga aggatgggac ggtctcggag ccggagctcg 240
tcccgctcca agcacaccaa gagcagcaag cacaacaaga agcgcagccg gtcccggctg 300
cgatcccggg acaaggagcg cgtgcggaag cgttccaaat ctccgggaaag taaacggaac 360
cggcgcgagg agtcgcgggc ccgttcgcgc tccaccaaca cggccgtgtc ccggcgcgag 420
cgggaccggg agcgcgcctc gtccccgccg gaccgcacgc acatcttcgg gcgcacgggtg 480
agcaagcgca gcagcctgga cgagaagcag aagcgagagg aggaggagaa gaaagcggag 540
ttcgagcggc agcgaaaaat tcgacagcaa gaaatagaag aaaaactcat cgaggaagaa 600
acagcacgaa gagtagaaga attggtagca aaaagggtgg aggaagaact ggagaaaagg 660
aaggatgaaa ttgaacgaga agttctccga aggggtggagg aagccaaacg catcatggaa 720
aagcagttgc tcgaagaact cgagcgacag agacaagctg agcttgccgc acaaaaagct 780
agagaggagg aagaacgtgc aaaacgtgag gagctagagc gaatactgga agagaataac 840
cgaaaaattg cagaagcaca agccaaactg gccgaagaac agttgagaat tgttgagaa 900
caaagaaaaa ttcatgagga aaggatgaaa ctagaacaag aacgacaacg tcaacaaaaa 960
gaagaacaaa aaattatcct gggcaaggcg aagttccagg caaaactgtc cttctcatta 1020
aaaaccagg attaaattgc aaactctgaa ctttttacia agaaaaatgg aaaaactttg 1080
tatggtagct tcatgttgaa gtggtttttt gtttttggtt ttgttt 1126

```

&lt;210&gt; 34

&lt;211&gt; 490

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2197211CB1

&lt;400&gt; 34

```

gcctctgaga aaagaaggtt ggaattatcg taatttgttt ctaggctgag ataccagcat 60
ggagaaaaatg ttggagtgtg cattcatagt cttgtggctt cagcttggtt ggttgagtgg 120
agaagaccag gtgacgcaga gtcccaggc cctgagactc caggaggagg agagtagcag 180
tctcaactgc agttacacag tcagcggttt aagagggtcg ttctgggtata ggcaagatcc 240
tggaagaggc cctgaattcc tcttcaccct gtattcagct ggggaagaaa aggagaaaga 300
aaggctaaaa gccacattga caaagaagga aagctttctg cacatcacag cccctaaacc 360
tgaagactca gccagttatc tctgtgctgt gcaggggggc attgggaatg tgctgcattg 420
cgggtccggc actcaagtgg ttgttttacc acatatccgg gaccctggac ctgccgtgta 480
acagctgaga

```

<210> 35  
 <211> 1799  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2263653CB1

<400> 35  
 ctaggccggt ggggtgagtgc accgcgttct cgcacgcgtc atggcggtcc tcggagtaca 60  
 gctgggtggtg accctgctca ctgccaccct catgcacagg ctggcgccac actgctcctt 120  
 cgcgcgctgg ctgctctgta acggcagttt gttccgatac aagcaccctg ctgaggagga 180  
 gcttcggggc ctggcgggga agccgaggcc cagaggcagg aaagagcggt gggccaatgg 240  
 ccttagtgag gagaagccac tgtctgtgcc ccgagatgcc ccgttccagc tggagacctg 300  
 cccctcacg accgtggatg ccctggctct gcgcttcttc ctggagtacc agtggtttgt 360  
 ggaactttgt gtgtactcgg gcggcggtgta cctcttcaca gaggcctact actacatgct 420  
 gggaccagcc aaggagacta acattgctgt gttctgggtg ctgctcacgg tgaccttctc 480  
 catcaagatg ttcctgacag tgacacggct gtacttcagc gccgaggagg ggggtgagcg 540  
 ctctgtctgc ctcacctttg ccttctctct cctgctgctg gccatgctgg tgcaagtggg 600  
 gcgggaggag accctcgagc tgggcctgga gcctggctcg gccagcatga cccagaactt 660  
 agagccactt ctgaagaagc agggctggga ctgggcgctt cctgtggcca agctggctat 720  
 ccgcgtggga ctggcagtggt tgggctctgt gctgggtgcc tctctacct tcccaggcct 780  
 gcggctggcc cagaccacc gccagcact gaccatgtcg gaggacagac ccatgctgca 840  
 gttctctctg cacaccagct tctgtctctc cctgttctac ctgtggctct ggacaaagcc 900  
 cattgcacgg gacttctctg accagccgcc gtttggggag acgcgtttct cctgctgtc 960  
 cgattctgcc ttcgactctg ggccctctg gttgctgggt gtgctgtgcc tgctgcggct 1020  
 ggcggtgacc cggccccacc tgcaggccta cctgtgctg gccaaaggcc ggggtggagca 1080  
 gctgcgaagg gaggtggcc gcacgaagc ccgtgaaatc cagcagaggg tgggtccgagt 1140  
 ctactgctat gtgaccgtgg tgagcttgca gtacctgac ccgctcatcc tcacctcaa 1200  
 ctgcacactt ctgctcaaga cgctgggagg ctattctctg gccctgggcc cagctctct 1260  
 actatcccc gacctatct cagccagcgc tgcccccatc ggctctgggg aggacgaagt 1320  
 ccagcagact gcagcgcgga ttgccggggc tctgggtggc ctgcttactc cctcttct 1380  
 ccgtggcgct ctggcctacc tcatctggtg gacggctgcc tgccagctgc tcgccagctc 1440  
 tttcggcttc tacttccacc agcacttggc aggcctctag ctgcctgcag accctctgg 1500  
 ggccctgagg tctgttctct gggcagcggg acaactagcct gccccctctg tttgcgcccc 1560  
 cgtgtcccca gctgcaaggt ggggcccggac tccccggcgt tcccttcacc acagtgcctg 1620  
 acccgcggcc ccccttgagc gccgagtttc tgccctcaga ctgtctctcc tgggcccagc 1680  
 agcatgaggg tcccagggcc attgtctccg aagcgtatgt gccaggtttg agtggcgagg 1740  
 gtgatgctgg ctgctcttct gaacaaataa aggagcatgc cgatttttaa aaaaaaaaaa 1799

<210> 36  
 <211> 1435  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2504590CB1

<400> 36  
 catagcgca ggttcacat gttggccagg ctgggtcttga actcctgtac cttgtgaaat 60  
 gcctgcctca gcctcccaaa gtgctgggat tacaggcatg agccactatg ccaggctca 120  
 aagatccttt ttttgttat cagatggagt ctactgtgt cccagagctg gattgcagtg 180  
 gcgcgacttc aactcactgc aagctctgcc tccgggttc acgccattct cctgctcag 240  
 gctcccgagt agctgggact acaggcgccc gccaccactc ctggctaatt ttgtttgtat 300  
 tttcagtaga gacgggatat caccatgtta gccaggatgg tctcgatctg cctgacctcg 360  
 tgatccgccc tctcagctc cccaaagtat taggattaca ggcgtgagcc accgcccccg 420  
 gccctcaaag attgttatga gtcctcacag atgtcagagg acgatgctgt gacctatatt 480  
 tgcaaacgcc cggaagactt aaagctattg ttttgcaaat gcacagataa caataataac 540  
 atgatgctgt ggtgagcat gaaagtttt ctctactcc taattataat gtttccctc 600  
 ttactgcttt gccctgctga tgggaattcag taaaggaatc tctaccagca gacacttgg 660  
 ttttaccctt cctaggacct ttaatatagata tctttttatt actaatcttt gctctttgct 720  
 tgtttaacct ccttgtaaag tttgtgtctt ccagattaca ataattccat gtaaagatga 780

```

tgctggcaca aggctttcaa cccatcccct cttctgacce agaagataaa gacatcctac 840
ctttgagcct tttagaacag gtatccaggg attttacctc tccagtgcct ggcagggtct 900
atgcccataa catcagcagg aagcagttac agaagatgaa cctccgccct tctgcaagcc 960
ccttaagatt aaggaggagt atataatctc tgatggggaa atgaggtagg agaccagaag 1020
gacttatttt ccattcccaa cccatttgaa cagagcagga tctggtcaaa acaggggtgca 1080
gtggagaagc ctgctgaaac cagcagatga tgatgaaagt gacctctagt tgcctcact 1140
gcttatgagc ataaagacac taccactggg accatggcca gtttacaat gtcattggca 1200
cacaccttgg caatggcctg gaagttactt tatatggttc tggaaactcc ctgccccttt 1260
cccagaaagt tctgaataac ctacctctta attggcatgc aattaaaagt ggctctaaat 1320
acaactagct agtagccac aggcaccaac tctgggcaca ctgcctatag gttagccctg 1380
ctctgcaaga agtagcacca gttcaataaa agttgctttc tctcaaaaaa aaaaa 1435

```

&lt;210&gt; 37

&lt;211&gt; 1792

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 2529619CB1

&lt;400&gt; 37

```

acacgcgtcc ggagctaacc aggcacatct tcacaatgaa atcttcacaa gctcagaacc 60
atcaacttgc tactctgtgt tcctccagtg cccagttcac aggaagcact caagagtagg 120
atcaattgtt atcaacctct agcagattat tgaatagacg gtcgttacat gaaactgaca 180
ctcagttctt catggtccac gatagttctt atatccagaa aacgtactta cgttatcaac 240
tctgactcca taagacaact ttgcaaagtc ttggctcata actataaatc aaggctgaat 300
caaacaggac tactgttctg aaagaatact cagactatga gtagtgaaat ggtgaaaaat 360
cagacaatgg tcacagagtt cctcctactg ggatttctcc tgggcccaag gattcagatg 420
ctcctctttg ggctcttctc cctgttctat gtcttcaccc tgctggggaa tgggaccatc 480
ctggggctca tctcactgga ctccagactc cacaccccca tgtacttctt cctctcacac 540
ctggccgtcg tcaacatcgc ctatgcctgc aacacagtcg cccagatgct ggtgaacctc 600
ctgcatccag ccaagcccat ctccctttgt ggctgcata caactacctt tctctttttg 660
agttttgcac atactgaatg cctcctgttg gtgctgatgt cctacgatcg gtacgtggcc 720
atctgccacc ctctccgata ttcatcacc atgacctgga aagtctgcat cactctggcc 780
atcactcctt ggacatgtgg ctccctcctg gctatgggtc atgtgagcct catectaaga 840
ctgccccttt gtgggcctcg tgaaatcaac cacttcttct gtgaaatcct gtctgtcctc 900
aggctggcct gtgctgatac ctggctcaac caggtgggtc tctttgcagc ctgcatgttc 960
atcctgggtg gaccactctg cctgggtgct gtctcctact cacacatcct ggccggccatc 1020
ctgaggatcc agtctgggga gggccgcaga aaggccttct ccacctgctc ctcccacctc 1080
tgcgtagtgg gactcttctt tggcagcgcc atcgtcatgt acatggcccc taagtcccg 1140
catcctgagg agcagcagaa ggtccttttt ctattttaca gttctttcaa cccgatgcta 1200
aaccccccta tttacaacct gaggaatgta gaggtcaggt gccctgagga gagcactgtg 1260
caagaaaagt cattcctaag aggtgtgaca tttgaactgc cagcctcagt tgtcacgtgg 1320
actcttgatg cccaattatt gcctcaatcc agaaaagttt acttctcttt atctgtgctt 1380
tactgacaga agggcaagtc ttctctcggt ttttgcatg aaaatttttag atgtgttgca 1440
ttcattgggt ttctatgaga tgtggtttta tcagacaatt ttttctttta tttcacaatt 1500
actttaatat ctttgaccag caaatatctc caccctccag ggagaggtag tagcttctag 1560
ggaaaccatc ttggagaggg tcctgtcttc ccctgaggtg ggctctgaat ccagcactct 1620
tcccctttct tggagggtca cttggaacca gtaactttt cagggctcct tcttcttagt 1680
tctgtcata catctgtcat gtaacacttt agtgttctat atttgcatag ctgtatcctt 1740
ccattagttt gtatagagct gacttagtat ttttcggacg cagttggctg at 1792

```

&lt;210&gt; 38

&lt;211&gt; 1817

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5467661CB1

&lt;400&gt; 38

```

gacggtccgg ccgcgcccca ggccgggagg tggagcgagg ctgccgccag ctggtccact 60

```

```

gacaactgct gccagctgga aggcgcgtag ggcgagaggt ggcgcgtcct cagctttcca 120
gtcttcgatt ccgcccgtt tccgagttag aaacgcaggc ctctctcttc cgtgaagcag 180
cgtcttgctc ctgcacctcc tctgtgctgt ctctttaaga gggctggctg ctggcgggat 240
ggacaccctg gaggaggtga cttggggcaa tgggagcaca gcgctacccc caccctggc 300
accaaaccatc agtgtgcctc atcgtgcctc gctgtgtctc tacgaagaca ttggcacctc 360
cagggctccg tactgggacc tctgtgtgct catccccaat gtgctcttcc tcatcttctc 420
gctctggaag ctcccatctg ctcgggcgaa gatccgcac accctccagcc ccatttttat 480
caccttctac atcctggtgt ttgtggtggc gctggtgggc attgcccggg ccgtggtatc 540
catgacggtg agcacctcga acgctgcaac tgttgctgat aagatcctgt gggagatcac 600
ccgcttcttc ctgctggcca tcgagctgag tgtgatcatc ctgggcctgg cctttggcca 660
cctggagagt aagtccagca tcaagcgggt gctggccatc accacagtgc tgtccctggc 720
ctactctgtc acccagggga ccctggagat cctgtaccct gatgcccac tctcagctga 780
ggactttaat atctatggcc atggggggcc ccagttctgg ctgggtcagct cctgcttctt 840
cttctgggtc tactctctgg tggctcatct tcccaagacc ccgctgaagg agcgcatctc 900
cctgccttct cggaggagct tctacgtgta tgcgggcac ctggcactgc tcaacctact 960
gcaggggctg gggagtgtgc tgtgtgctt cgacatcatc gaggggctct gctgtgtaga 1020
tgccacaacc ttcctgtact tcagcttctt cgctccgctc atctacgtgg ctttctctcg 1080
gggcttcttc ggctcggagc ccaagatcct cttctcttac aaatgccaa tggacgagac 1140
agaggagcca gatgtacacc taccacagcc ctacgtgtg gcccgacggg agggcctgga 1200
ggctgcaggg gctgtctggg cctcagctgc cagctactcg agcacgcagt tcgactctgc 1260
cggcggggtg gcctacctgg atgacatcgc ttccatgccc tgccacactg gcagcatcaa 1320
cagcacagac agcgagcgct ggaaggccat caatgcctga gggcagctgc cagggcctgt 1380
ggaggacagg ccagagagga ggccagcagg ccagagctcc ccaggggagg aggaccaggt 1440
caagggacgt tctgtgggca gtagccctgt gtggccctgt tcccaccatg agtctggagg 1500
ccccacctcc ctggggctcc caatccctt tgccatctct gctctcactg gggacctctc 1560
tccccttccc acctgtcttc atactgtcga gtgacatggc ccaggctttc cttccagggc 1620
catgcttggc aaggttggct gagggcaccc tccttctctg cacccttggc acgagggcag 1680
ggctggctct cccaatgcct ccaccccatc ccatgggtgc tttggcctcc tcaaagcatc 1740
caccatggtg gatggactga agtgtgtata ttttcttgat ctatttttta ataaaaagga 1800
aaaggagcaa aaaaaaa 1817

```

&lt;210&gt; 39

&lt;211&gt; 1820

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 229740CB1

&lt;400&gt; 39

```

ctttgagatg ctgaagatga tggaaagtcta taaggaaacct agagaacagc ccatcttcac 60
caccgagcgc catgtcttcc agattgaccc caacaccaag aagaactgga tgcctgcgag 120
caagcaggcg gtcaccgttt cctacttcta tgatgtcaca aggaacagct atcggtatcat 180
cagtgtggac ggagccaagg tgatcataaa cagcacatc acaccgaata tgaccttcac 240
caacacgtca cagacgtctg ggcagtgggc cgacagcaga gccaaacacag tgtttgggtt 300
ggggttttcc tctgagcagc agctgacaaa gtttgagag aaattccagg aggtgaaaga 360
agctgccaaag atagccaaag acaagacgca ggagaaaatc gagacctcaa gtaatcattc 420
ccaagcatcc agtgtcaacg ggacggacga tgaaaaggcc tctcacgccc gtccagccaa 480
cacacacctg aagtctgaga atgacaagct gaagattgcc ttgacgcaga gcgcagccaa 540
cgtgaagaag tgggagatcg agctgcagac ctttcgggag agcaatgcac ggctgaccac 600
agcactgcag gagtccggcag ccagtgtgga gcagtggaa aggagttct ccatctgcca 660
tgatgagaat gaccagctcc gcaacaagat tgatgagctg gaagaacaat gcagtgagat 720
caacagagag aaggagaaga acacgcagct gaagaggagg atcgaggagc tggaggcaga 780
gctccgagaa aaggagacag agctgaaaga tctccgaaaa caaagtgaat tcatacctca 840
gctcatgtca gagtgcgaat atgtctctga gaagctagag gcggcagaga gagacaatca 900
aaacctgga gacaaaagtgc gttccttaaa gacagacatt gaggagagca aataccgaca 960
gcgccacctg aaggtggagt tgaagagctt cctggagggtg ctggacggga agattgacga 1020
cctgcatgac ttccgcccag ggctctccaa gctgggcacc gataactagg gctggccgag 1080
gcccaggccc cgcccgtgag tcccagcgt gtgtgcgaga ccagatagct ctaggacgtt 1140
cttctgtgtg cattgtcttct gtaaattgcag gcgcagtttg tctgttttcc aaaccagttg 1200
tgccgtccac tcaactcttt tcagaataga aatctctct cgcttctctg gccttgtgag 1260
gttgtggaca actggaagat tctgactcag gaatccagaa ctaggctctac cttcaacatt 1320
tatgcagtca gggcagggat gtttatatct ttcataaggg ctgttgcaac catatgaact 1380

```

```

gaaaaaacac gcatttttcta atccaaatat tgatattctt tacaccaagc catcaggctc 1440
cttttatcaa atagcattca gaggatttga atgtccacca gacaccagcc ccgggggggca 1500
cagagagAAC aacattcctc tctgtcaaca tCGagaggct ttaaaacaac tgttttagtgg 1560
aaactttctg agagatggaa aacaagcttc tgggtgggtgc attttctggc cgggagttgc 1620
ctgcatccac gctactgccc cctgcccccc gcccccccag tttgtacggt tgcaacagtg 1680
ttccttttct tgggttttaat ttctgagcag atgattgtgc tgtgggaaca gcacacagtg 1740
aggggtgcta gcacaatgtc tggcacaag taggtgctta ataaatattt gttcaattaa 1800
atgtgaaaaa aaaaaaaaaa 1820

```

<210> 40

<211> 1620

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1317467CB1

<400> 40

```

tccccgccat gtgacgccgt ccttagccct gcgaccccca gcgcgtcccg ggccctgcgcc 60
tccgccccgc cgcgcagcac gatgcttctg ccgggacgcg cacgccaacc gccgacgcc 120
cagcccgtgc agcatcccg gctcgcggg caggtagagc cgccggggca gctcctgcgc 180
ctcttctact gcactgtcct ggtctgtctc aaagagatct cagcgtcac cgacttctct 240
ggttacctaa ccaaactcct gcaaaaccac accacctatg cctgtgatgg ggactatttg 300
aatctacagt gccctcggca ttctacgata agtgtccaat cggcatttta tgggcaagat 360
taccaaatgt gtagttccca gaagcctgcc tcccagaggg aagacagctt aacctgtgtg 420
gcagccacca ccttcagaa ggtgctggac gaatgccaga accagcgggc ctgccacctc 480
ctggtcaata gccgtgtttt tggacctgac ctttgtccag gaagcagtaa atacctcta 540
gtctccttta aatgccaacc taatgaatta aaaaacaaaa ccgtgtgtga agaccaggag 600
ctgaaactgc actgccatga atccaagttc ctcaacatct actctgcgac ctacggcagg 660
aggaccagg aaagggacat ctgctcctcc aaggcagagc ggctcccccc ttctgattgc 720
ttgtcttact cagctttgca agtctatccc cgaagggtgt atgggaagca gagatgcaa 780
atcatcgtca acaatcacca ttttggaaag cctgttttgc caggcgtgaa aaaataacct 840
actgtgacct acgcatgtgt tcccaagaac atactcacag cgattgatcc agccattgct 900
aatctaaaac cttctttgaa gcagaaagat ggtgaatatg gtataaactt cgaccaagc 960
ggatcgaagg ttctgaggaa agatggaatt cttgttagca actctctggc agcctttgct 1020
tacattagag cccaccggga gagagctgcc ctgctgttcg tgtccagtgt ctgcatcggc 1080
ctggccctca cactgtgcgc cctggtcac agagagtcct gtgccaagga cttccgcgac 1140
ttgcagctgg ggaggaggca gctggtgcca ggaagtgaac aggtcgagga ggacagcgag 1200
gatgaagaag aggaggagga cccctctgag tctgatttcc caggggaaact gtcgggggtc 1260
tgtaggactt catatcctat atacagtccc atagaagctg cagagctcgc agaaaggatt 1320
gagcgcaggg agcaaatcat tcaggaaata tggatgaaca gtggtttgga cacctcgtc 1380
ccaagaaaca tgggcccagt ctactgaaaa ccacatgcat cttgatgca tcgactttc 1440
tgaagaagga aggatcccaa atgccccctc agttctggtt cacctgtacc ttctatgaag 1500
gagaattcgt catgtcattc aacactcgtg aggccaggaa gctattaaag ggatgtttca 1560
agctgtttct agcacattcc aaaataaatg aggagggaag agtctttgtt ttctgtaaaa 1620

```

<210> 41

<211> 974

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2279267CB1

<400> 41

```

atgggggaca atataacatc catcacagag ttctctctac tgggatttcc cgttggccca 60
aggattcaga tgctcctctt tgggtctctt tccctgttct acgtcttcac cctgctgggg 120
aacgggacca tactggggct catctcactg gactccagac tgcacgcccc catgtacttc 180
ttctctctac acctggcggt cgtcgacatc gcctacgctt gcaacacggt gccccggatg 240
ctgggtgaacc tctgtcatcc agccaagccc atctcctttg cgggcccgcg gatgcagacc 300
tttctgtttt ccacttttgc tgtcacagaa tgtctcctcc tgggtggtgat gtcctatgat 360
ctgtacgtgg ccactctgca cccctccga tatttggcca tcatgacctg gagagtctgc 420

```

```

atcacccctcg cgggtgacttc ctggaccact ggagtccttt tatccttcat tcatcttctg 480
ttactttctac cttttaccctt ctgtaggccc cagaaaaattt atcacttttt ttgtgaaatc 540
ttggctgttc tcaaacttgc ctgtgcagat acccacatca atgagaacat ggtcttggcc 600
ggagcaattt ctgggctggg gggacccttg tccacaattg tagtttcata tatgtgcac 660
ctctgtgcta tccttcagat ccaatcaagg gaagttcaga ggaaagcctt ctgcacctgc 720
ttctcccacc tctgtgtgat tggactcttt tatggcacag ccattatcat gtatgttgg 780
cccagatatg ggaaccccaa ggagcagaag aaatatctcc tgctgtttca cagcctcttt 840
aatcccatgc tcaatcccct tatctgtagt cttaggaact cagaagtga gaatactttg 900
aagagagtgc tgggagtga aagggttta tgaaaaggat tatggcattg tgactgacag 960
tgacctagga agtt 974

```

<210> 42

<211> 1561

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2436258CB1

<400> 42

```

gttaaaacag ggatgtgcag atggaggtcg gaggagacac tgctgccccg gccccggggg 60
gcgcggagga cttggaggac acgcagttcc ccagtgagga agctagagaa ggtggagggg 120
ttcacgcggt cccgccgat cccgaagacg agggcctgga ggaaacagag gaccacaagc 180
tgggtgttcc acaacagggc ccaactgttg tctgtggccat gtcacggact tctcagtcag 240
cagcccagct gcggggggag ctgctagctg tgcacgcaca gatcgtgagc acacttacac 300
gtgcaagtgt cgcgcgcac ttcgcacaca agcagaacta tgacctccgc cgcctgctgg 360
ctggttcaga gcgcacactg gaccgacttc tggacagtat ggagcaggac ccaggagccc 420
tgctcctggg tgccgtgcgc tgtgtgcccc ttgccccccc gctgcgagac gcactaggtg 480
cgctcctccg acgttgca ca gcgcctggcc tggcgctgtc agtgctggca gtaggcggtc 540
gacttataac agcagcccag gagcggaatg tgctggccga gtgccggctg gaccagctg 600
acctgcagtt gctgctcgac tgggtgggtg caccagcctt tgcggcggtg gaggtctggg 660
cacctgtgtg cctgccccgc tccaacctg atggttttt ctacgcctac gtggccccgc 720
tggatgctat gcctgtctgc ctgctgctgc ttggcaccca acgtgaagcc ttccatgcca 780
tggccgctg cggcgccctg gttgaagatg ggatgcatgc ccttgggtgc atgcgtgccc 840
ttggggaggc tgccagcttc tctaagtcc catcagccag tgctcctgcc tacagcgtgc 900
aggctgtcgg ggcgcggggc ctccggcact tctgtataa gccgctggac atccctgacc 960
accaccgcca actgccccag tttaccagcc ctgagctaga ggccccctac agcagagagg 1020
aggagcggca gcggctgtcg gacctgtacc accgcctgca tgctcgtctc cacagcacct 1080
cccgaccct gcgcctcatt taccacgtgg ctgagaagga gacactactg gcctgggtga 1140
cctccaaatt cgagctctat acctgcctca gccctctggt gaccaaggca ggtgcaatct 1200
tggtagtgac caaactcctg cgctgggtga agaaagagga ggaccggctc ttcattcggt 1260
accacccaa gtactccaca ccaccagcca cctctacgga ccaagctgcc cataatggct 1320
tgttcactgg actctgatag ttggagctcc cagaccaggc agtgctggga gcaaccacct 1380
ttgtttttta ccttctgtct accctggaaa tgtgtgtggg ggtgtgtctg tggccagtca 1440
ttgtctccct aagcaatggg gcaaggtctg agggcccacc gatgagagag atggtggcag 1500
ccgacagggg agcaggctgc tttccctgcc cagtcatgca ccttcccctc tggggaaatt 1560
c 1561

```

<210> 43

<211> 1619

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2681738CB1

<400> 43

```

tgcgagggcg ggatagctgt ccaaggtctc ccccagcact gaggagctcg cctgctgccc 60
tcttgccgcg ggaagcagc accaagttca cggccaacgc cttggcacta ggtccagaa 120
tggctacaac agtccctgat ggttgccgca atggcctgaa atccaagtac tacagacttt 180
gtgataaggc tgaagcttgg ggcacgtccc tagaaacggt ggccacagcc ggggttgtga 240
cctcggtggc cttcatgctc actctcccga tctcgtctg caaggtgcag gactccaaca 300

```

```

ggcgaaaaat gctgcctact cagtttctct tcttctctggg tgtgttgggc atctttggcc 360
tcaccttcgc cttcatcatc ggactggacg ggagcacagg gccacacgc ttcttctct 420
ttgggatcct cttttccatc tgcttctcct gctgtctggc tcatgctgtc agtctgacca 480
agctcgctcg ggggaggaag cccctttccc tgttgggtgat tctgggtctg gccgtgggct 540
tcagcctagt ccaggatggt atcgctattg aatatattgt cctgaccatg aataggacca 600
acgtcaatgt cttttctgag ctttccgctc ctgctcgcaa tgaagacttt gtctctctgc 660
tcacctacgt cctcttcttg atggcgctga ctttctcat gtcctccttc acctctctgt 720
gttcttccac gggctggaag agacatgggg cccacatcta cctcacgatg ctctctcca 780
ttgccatctg ggtggcctgg atcaccctgc tcatgcttcc tgactttgac cgcaggtggg 840
atgacaccat cctcagctcc gccttggctg ccaatggctg ggtgttctctg ttggcttatg 900
ttagtcccga gttttggctg ctcacaaagc aacgaaaccc catggattat cctgttgagg 960
atgctttctg taaacctcaa ctctggaaga agagctatgg tgtggagaac agagcctact 1020
ctcaagagga aatcactcaa ggttttgaag agacagggga cacgctctat gccccctatt 1080
ccacacattt tcagctgcag aaccagcctc cccaaaagga attctccatc ccacgggccc 1140
acgcttggcc gagcccttac aaagactatg aagtaaaaga agagggcagc taactctgtc 1200
ctgaagagtg gtagcaaatgc agccgggcgg cagatctagc gggagctcaa agggatgttg 1260
gcgaaatctt gagtcttctg agaaaactgt acaagacact acgggaacag tttgcctccc 1320
tcccagcctc aaccacaatt cttccatgct ggggctgatg tgggctagta agactccagt 1380
tcttagaggc gctgtagtat tttttttttt ttttggctca tcttaggat acttctttta 1440
agtgggagtc tcaggcaact caagtttaga cccttactct ttttgtttgt tttttgaaac 1500
aggatcttgc tctgtcacc caggttgagt gcagtggtgc gatcacagcc cagtgcagcc 1560
tcgaccacct gtgctcaagc aatcctccca tctccatctc ccaaagtgtc gggatgaca 1619

```

<210> 44

<211> 3691

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2859482CB1

<400> 44

```

ggcgagcggc gggcgggcga ggaggagacg gcaggctcga ctgaccaa atggaagtgt 60
gcaagtcttg aacacattag ttatataaat gtggcctgaa agtttccctc ctctttccaa 120
atcaaccata atattctgag gcattaaaaa atatttaatc attcatgtgt tgagactcat 180
tcttgagtta tggatgacaa ggcttctgtt ggaaaaatca gtgtctcttc agactcagta 240
tctactctta atagtgaaga ttttgtcttg gtttccaggc aaggagatga gacaccatct 300
acaaataatg gaagtgatga tgagaaaaca ggactcaaga ttgtagggaa tgggaagtga 360
cagcagctgc aaaaagagct agcagatgta ctgatggatc ctccaatgga cgaccagcca 420
ggggaaaaag agcttgtgaa aaggtcacaa ctggatgggt aaggagatgg gcctctttct 480
aatcagctct ccgcttcatc caccattaac cctgtgccat tagtagggct ccaaaaaacca 540
gagatgagcc taccagtga acctggacaa ggagattctg aagcttcaag tcttttcaca 600
ccagtggccg atgaggacag cgtagttttc agtaaaactga cttacttagg ctgtgcctcg 660
gtaaatgtct ccaggagtga agtggaagcc ttaaggatga tgtccatctt aagaagccag 720
tgtcagattt cactagatgt taccctttca gtgccgaatg tgtctgaagg aattgtgaga 780
ctcttagatc ctacagacaa cactgaaata gcaaactacc ctatctacaa aatcctcttc 840
tgtgtcagag ggcatgatgg aactcctgag agtgactgtt ttgctttcac tgaaagtcac 900
tacaatgcag agctcttcag aatacacgtc ttccgggtgt aaatacaaga agctgtaagc 960
cggatacttt acagttttgc cactgccttc cgcggttctg ccaagcagac cccactttca 1020
gccactgctg caccacagac tcttgacagt gacatcttta ccttctctgt gtcttttagaa 1080
ataaaagaag atgatggtaa aggttatatt agtgacgttc ccaaagataa ggacagacag 1140
tgctttaaac tacgccaagg aattgataag aagattgtca tctatgtgca gcaaacaact 1200
aataaagaac ttgccattga aaggtgtttt ggtcttctcc ttagtccagg aaaagatgta 1260
cgaaatagtg acatgcactt attagatttg gaatctatgg gcaaaaagttc agatggaaag 1320
tcgtatgtta ttacggggag ctggaatcca aaatccccac attttcaagt tgtaaatgaa 1380
gaaactccta aagataaagt cctgtttatg accacagctg tagatttggt aataacagaa 1440
gtacaggagc ctgttcgatt tctcctggag acaaaaagtc gcgtttgctc acctaatagaa 1500
agattattct ggcccttcag caaacgtagt actactgaaa atttcttttt gaaactaaaa 1560
cagataaagc aaagggagag aaagaataat actgacactt tatatgaagt tgtatgcttg 1620
gaaagtgaat cagaaaagaga gaggaggaaa actacagcca gtccttcagt tcgcctgcca 1680
cagctgtgat cgaaaagttc agtgatacct tctcctccag aagatgatga agaggaagat 1740
aatgatgaac ctctctgtag tggatctggt gatgtatcca aagaatgtgc agaaaaaatt 1800
cttgaaacat ggggagaact gttgtcaaaa tggcatctca acttgaatgt gagaccgaag 1860

```

cagttgtcat	ccttagtaag	aaacggtgtc	cctgaagctc	ttcaggagaga	agtctggcag	1920
ctgctagcag	gctgtcataa	caatgaccac	ctggtagaga	aataccgcat	tcttatcaca	1980
aaggagtctc	cccaggacag	tgctatcacc	cgggatatta	accgaacatt	cccagcccat	2040
gactacttta	aggacacagg	aggagatgga	caagattcct	tatataaaat	atgcaaggct	2100
tattctgtgt	atgatgaaga	gattgggttat	tgccagggcc	agtcatttct	tgctgctgtg	2160
ctccttctcc	atatgcctga	agaacaggca	ttcagtggtc	tggtcaagat	catgtttgac	2220
tatgggctca	gggaactttt	caagcaaaac	ttcgaagatt	tgcatgcaa	attttaccag	2280
ttggagcgcc	tcatgcagga	atacattcct	gacctgtaca	accacttctt	ggatataagc	2340
cttgaagcac	acatgtatgc	ctcccagtg	tttcttactc	ttttcactgc	aaaattccct	2400
ctctacatgg	tcttccatat	catcgacctg	cttttatgtg	agggaataag	tgttatTTTT	2460
aatgtcgccc	ttggattatt	aaagacttct	aaagatgacc	tgctgttgac	agactttgaa	2520
ggtgccttga	agttcttttag	ggttcagctt	cctaagagat	accgctcaga	agaaaatgca	2580
aaaaaactaa	tggaatttagc	ctgcaacatg	aagattagtc	agaagaagtt	gaaaaataac	2640
gagaaagaat	atcacacccat	gagggaacag	caggcccgagc	aagaagaccc	catcgagcga	2700
tttgagcggg	agaataggcg	tctacaagaa	gctaacatga	ggttggaaca	ggaaaacgat	2760
gacttagccc	atgagctggg	gaccagcaag	attgcactac	ggaaggacct	ggataacgct	2820
gaggaaaagg	cagatgctct	gaataaggag	ctgctgatga	ccaaacagaa	gttgattgat	2880
gcagaagaag	agaaaagacg	gctggaagaa	gagtctgtc	agttaaaaga	aatgtgccgt	2940
cgggaaactcg	acaaggcaga	atctgagatt	aaaaaaaaaca	gttctatcat	tggtgactat	3000
aagcagattt	gttctcagtt	gagtgaagaa	ttggagaagc	agcagacagc	caataagggtg	3060
gaaattgaga	aaattcggca	aaaagtggat	gactgtgagc	ggtgccggga	atTTTTcaac	3120
aaagaagggc	gtgtaaaagg	cataagctca	accaaggagg	ttttagatga	ggacacggat	3180
gaagagaaaag	agacgctcaa	gaaccagctg	agagaaatgg	agctagaact	ggcacagacc	3240
aaactccagc	tggtggaggc	cgagtgtgaa	atacaggact	tggaacacca	tttagggctt	3300
gccctcaatg	aggtgcaggc	agccaagaag	acgtggttta	accgaacact	gagctccata	3360
aagacagcaa	ccgggggttca	agggaaagag	acttgctgag	agcagctgcc	gcctcccgac	3420
accttcagaa	aacacgacac	cttttgttgc	cttctttggc	cagatgtgtg	attctgtgac	3480
ttgtcccagg	accagaatgt	acctaagtca	gatccataga	cgcagtgttg	taggtcactg	3540
gaccagagct	tgtgaagcag	gcaacctctg	gggtaagact	actgatacta	acaggcctgc	3600
tagctcagcc	gacgctctgg	acactctaga	aatcactcct	cagtgtgacc	tcccaggcct	3660
cttccccgtg	tacgtcaaca	cctcaccag	c			3691



(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
1 February 2001 (01.02.2001)

PCT

(10) International Publication Number  
WO 01/07612 A3

(51) International Patent Classification<sup>7</sup>: C12N 15/12,  
C07K 14/72, 16/28, G01N 33/50, 33/566, C12Q 1/68,  
A61K 38/17, 31/00

(21) International Application Number: PCT/US00/20035

(22) International Filing Date: 21 July 2000 (21.07.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/145,232 21 July 1999 (21.07.1999) US  
60/158,578 7 October 1999 (07.10.1999) US  
60/165,192 12 November 1999 (12.11.1999) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 60/158,578 (CIP)  
Filed on 7 October 1999 (07.10.1999)  
US 60/165,192 (CIP)  
Filed on 12 November 1999 (12.11.1999)  
US 60/145,232 (CIP)  
Filed on 21 July 1999 (21.07.1999)

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): AU-YOUNG, Janice [US/US]; 233 Golden Eagle Lane, Brisbane, CA 94005 (US). BANDMAN, Olga [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). TANG, Y., Tom [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). AZIMZAI, Yalda [US/US]; 2045 Rock

Springs Drive, Hayward, CA 94545 (US). BURFORD, Neil [GB/US]; 1308 4th Avenue, San Francisco, CA 94122 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). LU, Dyung, Aina, M. [US/US]; 55 Park Belmont Place, San Jose, CA 95136 (US). HILLMAN, Jennifer, L. [US/US]; 230 Monroe Drive, #12, Mountain View, CA 94040 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94025 (US). LAL, Preeti [IN/US]; 2382 Lass Drive, Santa Clara, CA 95054 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— With international search report.

(88) Date of publication of the international search report:  
7 June 2001

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RECEPTORS AND ASSOCIATED PROTEINS

(57) Abstract: The invention provides human receptors and associated proteins (RECAP) and polynucleotides which identify and encode RECAP. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with expression of RECAP.



WO 01/07612 A3

## INTERNATIONAL SEARCH REPORT

Intern. Application No

PCT/US 00/20035

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/12 C07K14/72 C07K16/28 G01N33/50 G01N33/566  
 C12Q1/68 A61K38/17 A61K31/00

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K G01N C12Q A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

BIOSIS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online]            accession number U22015,            15 March 1995 (1995-03-15)            SEOL W. ET AL.: "Mus musculus X receptor            interacting protein (RIP110) mRNA."            XP002153300            abstract            &amp; SEOL W. ET AL.: "Isolation of proteins            that interact specifically with the            retinoid X receptor: Two novel orphan            receptors."            MOLECULAR ENDOCRINOLOGY,            vol. 9, no. 1, 1995, pages 72-85,            XP000651076            the whole document</p> <p style="text-align: center;">---</p> <p style="text-align: center;">-/--</p>	1,3-7,9, 11,12

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&amp;" document member of the same patent family

Date of the actual completion of the international search

20 November 2000

Date of mailing of the international search report

27.02.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
 Fax: (+31-70) 340-3016

Authorized officer

Mandl, B

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/20035

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL [Online]  accession number AI337112,  31 December 1998 (1998-12-31)  STRAUSBERG R.: "Retinoid X receptor  interacting protein."  XP002153301  abstract</p> <p>---</p>	1,3-7,9, 11,12
X	<p>WO 98 45437 A (GENETICS INST)  15 October 1998 (1998-10-15)  SEQ.ID.822</p> <p>---</p>	3,12
P,X	<p>DATABASE EMBL [Online]  accession number AF113538,  12 December 1999 (1999-12-12)  PENG Y. ET AL.: "Homo sapiens retinoid X  receptor interacting protein mRNA."  XP002153302  abstract</p> <p>-----</p>	1,3-7,9, 11,12

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US 00/20035

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
  
Although claim 18 is directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☒ Claims Nos.: 20,21,23,24  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:  
  
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Claims 1-28 (all partially)

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 20,21,23,24

Claims 20, 21, 23 and 24 refer to agonists and antagonists of the polypeptides without giving a true technical characterization. Moreover, no such specific compounds are defined in the application. In consequence, the scope of said claims is ambiguous and vague, and their subject-matter is not sufficiently disclosed and supported (Art. 5 and 6 PCT). No search can be carried out for such purely speculative claims whose wording is, in fact, a mere recitation of the results to be achieved.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

## FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1: Claims 1-28 (all partially)

An isolated polypeptide as represented by SEQ.ID.1, and variants, biologically active fragments and immunologic fragments thereof, an isolated polynucleotide encoding said polypeptide, variants and fragments; a cell transformed with said polynucleotide; a transgenic organism comprising said polynucleotide; a method for producing said polypeptide, variants and fragments; and isolated antibody specific for said polypeptide; methods for detecting said polynucleotide or said polypeptide; methods for detecting agonists, antagonists or modulators of the activity of said polypeptide; pharmaceutical compositions comprising said polypeptide, agonist or antagonist; a method of screening for compounds that specifically bind to said polypeptide; a method for screening a compound for effectiveness in altering expression of said polynucleotide; and a method for assessing toxicity of a test compound comprising a hybridization probe derived from said polynucleotide.

Inventions 2-20: Claims 1-28 (all partially)

Same as subject 1 but limited to one polypeptide sequence selected from SEQ.IDs. 2-12, 14-20 and 22, wherein invention 2 is limited to SEQ.ID.2, invention 3 is limited to SEQ.ID.3, ..... invention 12 is limited to SEQ.ID. 12, invention 13 is limited to SEQ.ID.14, ..... invention 19 is limited to SEQ.ID.20, and invention 20 is limited to SEQ.ID.22.

## INTERNATIONAL SEARCH REPORT

### Information on patent family members

International Application No

PCT/US 00/20035

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9845437 A	15-10-1998	AU 6956798 A EP 0973899 A	30-10-1998 26-01-2000
-----			